Robust prediction of resistance to trimethoprim in Staphylococcus aureus

Article (Accepted Version)

Robust prediction of resistance to trimethoprim in *Staphylococcus aureus*[*]

Philip W. Fowler[†1], Kevin Cole[2], N. Claire Gordon[1], Angela M. Kearns[3], Martin J. Llewelyn[2], Tim E. A. Peto[1], Derrick W. Crook[1], and A. Sarah Walker[‡1]

[1]Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Headley Way, Oxford, OX3 9DU, United Kingdom

[2]Department of Infectious Diseases and Microbiology, Royal Sussex County Hospital, Brighton, Brighton and Sussex Medical School, BN1 9PS, United Kingdom

[3]Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, Public Health England, Colindale, NW9 5EQ, United Kingdom

January 5, 2018

KEYWORDS: clinical microbiology, antibiotic susceptibility testing, free energy calculations, antimicrobial resistance

[†]Lead and Corresponding author. Email: philip.fowler@ndm.ox.ac.uk, Twitter: @philipwfowler

[‡]MJL, TEAP, DWC and ASW are all Senior Authors

Summary

The rise of antibiotic resistance threatens modern medicine; to combat it new diagnostic methods are required. Sequencing the whole genome of a pathogen offers the potential to accurately determine which antibiotics will be effective to treat a patient. A key limitation of this approach is that it cannot classify rare or previously unseen mutations. Here we demonstrate that alchemical free energy methods, a well-established class of methods from computational chemistry, can successfully predict whether mutations in *Staphylococcus aureus* dihydrofolate reductase confer resistance to trimethoprim. We also show that the method is quantitively accurate by calculating how much the most common resistance-conferring mutation, F99Y, reduces the binding free energy of trimethoprim and comparing predicted and experimentally-measured minimum inhibitory concentrations for seven different mutations. Finally, by considering up to 32 free energy calculations for each mutation, we estimate its specificity and sensitivity.

INTRODUCTION

Resistance of bacteria to the antibiotics used to treat them is a substantial and growing global threat to human health (Davies, 2013; World Economic Forum, 2013). Measures to counter the emergence of antibiotic resistance are restricted by the limitations of conventional diagnostic microbiology. This predominantly still relies on culture-based, phenotypic identification of bacteria followed by growth in the presence of different antibiotic concentrations to detect resistance. The process is labour intensive, takes days or even weeks depending on the growth rate of the organism in question, is expensive and open to subjective interpretation. Genetic approaches, particularly those based on sequencing the entire genome of a pathogen (Didelot et al., 2012; Köser et al., 2014), have the potential to be faster and cheaper. Inferring the phenotype of an infecting pathogen from whole-genome sequence data by considering known resistance genes or mutations has already been shown to be reasonably accurate for a range of pathogens (Gordon et al., 2014; Walker et al., 2015; Pankhurst et al., 2016; Bradley et al., 2015) and has recently been implemented in the U.K. for the routine diagnosis of *M. tuberculosis* infections

(Walker et al., 2017). New mutations, however, continually arise and a genetics-based clinical microbiology service therefore also needs to be able to *predict* the effect of novel mutations. In this paper we demonstrate that molecular-based computational chemistry methods can predict whether individual protein mutations confer resistance to an antibiotic.

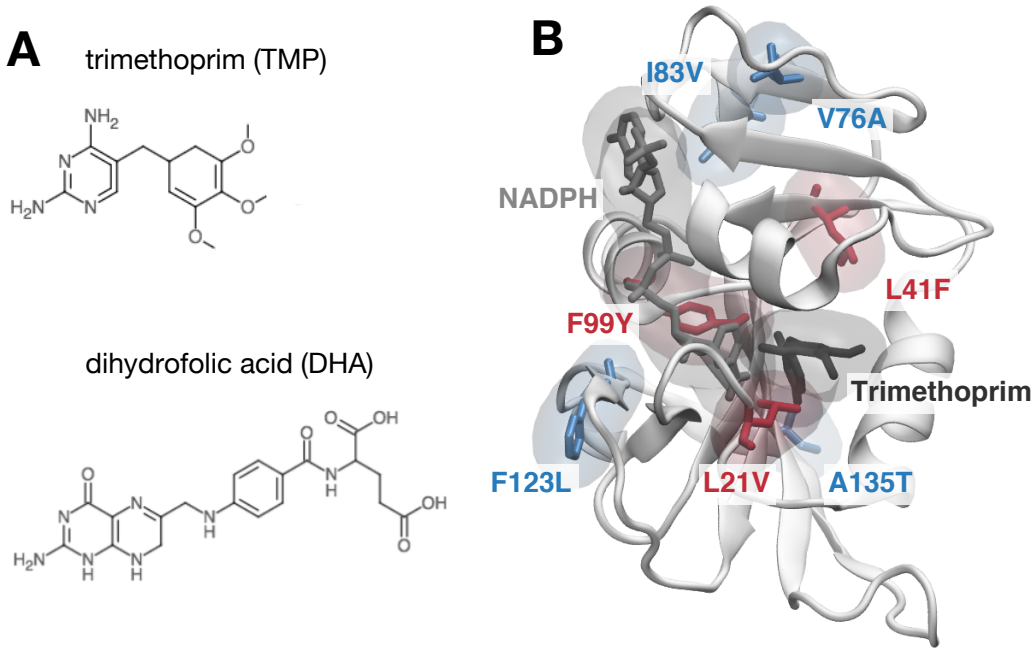

Figure 1: Seven mutations in *S. aureus* dihydrofolate reductase (DHFR) were chosen from a whole genome sequencing study of clinical isolates (Gordon et al., 2014) to test our approach (*A*) Trimethoprim (TMP) competes with the natural substrate, dihydrofolic acid (DHA), for binding to DHFR, thereby inhibiting the action of this essential protein. (*B*) A structure of chromosomal *S. aureus* DHFR (*dfrB*) bound with TMP and NADPH, as resolved by X-ray crystallography (Oefner et al., 2009). Three of the mutations, colored red (F99Y, F99Y/L21V & L41F), were previously shown to confer resistance to TMP, whilst the remaining four, colored blue (F123L, A135T, V76A, I83V), remained susceptible to the action of the antibiotic (Gordon et al., 2014). This classification was confirmed by independent measurement of TMP minimum inhibitory concentrations for each mutant (Table S1 & S2). These colors are used throughout.

As proof of principle we have investigated the effect of mutations to *Staphylococcus aureus* dihydrofolate reductase (DHFR) on the binding of the antibiotic trimethoprim (TMP, Fig. 1A). *S. aureus* is a clinically important gram-positive pathogen and has been the focus of much research due to the development of methicillin- and vancomycin-resistant strains, known as MRSA and VRSA, respectively. TMP, usually administered as co-trimoxazole (trimethoprim-sulfamethoxazole), has a long history of treating *S. aureus* infections (Tong et al., 2015) in-

cluding common skin and soft tissue infections caused by MRSA strains (Nurjadi et al., 2014). TMP competes with the natural substrate, dihydrofolic acid (DHA, Fig. 1A), for binding to DHFR, thereby preventing DHFR catalyzing the conversion of DHA to tetrahydrofolic acid. Since tetrahydrofolate is essential for the biosynthesis of thymidylate, purine nucleotides, and some amino acids, arresting the production of DHA inhibits bacterial growth. Resistance to TMP in *S. aureus* can either arise from mutations in the chromosomal gene *dfrB*, or from the introduction of other naturally-resistant genes (*dfrA*, *dfrG* and *dfrK*) via plasmids (Lowy, 2003; Nurjadi et al., 2014). Here we focus on seven mutations in the *dfrB* chromosomal gene. We have chosen this gene for five reasons: (i) a series of resistance-conferring and no-effect mutations have been identified via whole-genome sequencing of isolates from patient infections (Gordon et al., 2014), as well as by more traditional methods, (ii) the most common resistance-conferring mutation is a very small chemical change (Phe $\rightarrow$ Tyr) and this is therefore a challenging test for any predictive approach, (iii) DHFR is a small, soluble protein that has been well-studied, (iv) several experimental structures exist of *S. aureus* DHFR bound to TMP (Fig. 1B) (Dale et al., 1997; Oefner et al., 2009; Heaslet et al., 2009) and (v) there is published quantitative biophysical data on how the most common resistant-conferring mutation in *S. aureus* affects the binding of TMP to DHFR (Pires et al., 2015; Oefner et al., 2009; Dale et al., 1997; Frey et al., 2010, 2012). Since this is a classification problem we emphasise the importance of having *negative* controls (that is, mutations that are known to have no effect). This underscores the vital importance of clinical whole genome sequencing studies as these naturally identify large numbers of such mutations.

Our hypothesis is that chromosomal mutations in an open reading frame will confer resistance if the mutation causes the antibiotic molecule to bind less well to the encoded protein, whilst, crucially, not significantly affecting how well the natural substrate binds. This is only one of several mechanisms by which bacteria can evolve resistance to antibiotics (Blair et al., 2014). Other mechanisms include the introduction, by horizontal gene transfer, of genes encoding either proteins that degrade antibiotics, for example β-lactamases which are common in gram-negative bacteria, or, as mentioned above, naturally resistant versions of chromosomal proteins. The over-expression of efflux pumps can reduce the concentration of the antibiotic within the bacterium to below effective levels or the cell well can simply be impenetrable to

most antibiotics, the most notable example of this being *M. tuberculosis*.

The binding free energy (ΔG) is the thermodynamic quantity that captures how strongly a small molecule, like an antibiotic, is bound to a protein. Our hypothesis therefore distills down to calculating how a specific mutation affects the binding free energies, relative to the wildtype (*wt*), of both TMP and DHA

$$\begin{aligned} \Delta\Delta G_{TMP} &= \Delta G^{mutant}_{\text{TMP}} - \Delta G^{wt}_{\text{TMP}} \\ \Delta\Delta G_{DHA} &= \Delta G^{mutant}_{\text{DHA}} - \Delta G^{wt}_{\text{DHA}} \end{aligned}$$

Whilst it would be trivial for a mutation to disrupt the binding of the antibiotic (i.e. $\Delta G^{mutant}_{\text{TMP}} > \Delta G^{wt}_{\text{TMP}}$ and so $\Delta\Delta G_{\text{TMP}} > 0$), it is difficult for a mutation to simultaneously not disrupt the binding of the natural substrate (i.e. $\Delta G^{mutant}_{\text{DHA}} \sim \Delta G^{wt}_{\text{DHA}}$ leading to $\Delta\Delta G_{\text{DHA}} \sim 0$). For a mutation to give rise to a viable strain of *S. aureus* that is resistant to TMP a first estimate of a binding free-energy criterion is therefore is that $\Delta\Delta G_{\text{TMP}} > 0$ and $\Delta\Delta G_{\text{DHA}} \sim 0$. By making some simple assumptions and applying kinetic theory, we will relate these changes in binding free energies to the minimum inhibitory concentrations (MICs) of the antibiotics. This is the quantity measured by clinical microbiology laboratories, and we are able, through MICs distributions published by the European Committee on Antimicrobial Susceptibility Testing (EUCAST), derive more sophisticated criteria based on clinical data.

To calculate how the binding free energy of either the antibiotic or the natural substrate changes upon introduction of the mutation we will apply Hamiltonian-exchange thermodynamic integration, an alchemical free energy method (Fowler et al., 2005; Gilson and Zhou, 2007; Fowler et al., 2007; Michel et al., 2010; Chodera et al., 2011; Gapsys et al., 2015a; Perez et al., 2016; Abel et al., 2017). Alchemical free energy methods are derived from classical statistical mechanics and calculate the cost of perturbing a chemical moiety, such as an amino acid sidechain, into another using a series of classical molecular dynamics (MD) simulations; hence they are dubbed 'alchemical'. There are no free parameters, and so in theory are exact, although in practice there are always likely to be errors due to imperfections in the parametrisation of the molecules and the incomplete exploration of the dynamical phase space of the system during

the simulations. We will not consider here other methods of calculating or estimating binding free energies, such as computational docking, 'endpoint' methods or protein design or stability algorithms, since they are unlikely, in our opinion, to capture the subtlety of the molecular perturbations. Since each free energy calculated by an alchemical free energy method requires a number of molecular dynamics simulations, this approach potentially requires large amounts of computational resource; however, given the continued increase in computing speeds this class of methods is coming of age and is beginning to find application (Wang et al., 2015; Samsudin et al., 2016; Gapsys et al., 2016; Lenselink et al., 2016).

Traditionally, a single calculation would be run for each perturbation (here a protein mutation) and the error in the free energy estimated by, e.g. dividing the simulation trajectories into 'independent' sections by calculating a correlation time. Since this is a clinically important problem where the accuracy of the classification, and potentially also minimising the time taken to return a prediction, are essential, we shall instead run a large ensemble of relatively-short thermodynamic integration calculations for each mutation, simplifying the estimation of confidence intervals, as well as, subject to having sufficient computational resource, potentially reducing the time to solution. In the second half of the paper we will estimate the sensitivity and specificity of our method.

RESULTS

Clinically a mutation is described as resistant if the minimum concentration of an antibiotic that inhibits the growth of the bacteria is greater than a reference concentration. According to the European Committee on Antimicrobial Susceptibility Testing (EUCAST, 2016), *S. aureus* is defined as not susceptible to TMP (i.e. resistant) if its minimum inhibitory concentration (MIC) is ≥ 4 mg/l. Since TMP is a competitive inhibitor of DHFR and, assuming Michaelis-Menten enzyme kinetics (Price et al., 2009), then as shown in the Supplemental Information, if we assume that the mutation only affects the dissociation equilibrium constant of the antibiotic (K_i) we can derive a simple binding free-energy based resistance criterion,

$$\Delta\Delta G_{\mathrm{TMP}} \geq 0.8 \text{ kcal/mol}. \tag{R1}$$

This assumes that the enzyme rate constant and the concentrations of the enzyme and the substrate are all unaffected by the mutation. Alternatively, if we allow the protein mutation to affect the dissociation constants of *both* the inhibitor and the natural substrate, then we find a second resistance criterion,

$$\Delta\Delta G_{\mathrm{TMP}} - \Delta\Delta G_{\mathrm{DHA}} \geq 0.8 \text{ kcal/mol}. \tag{R2}$$

This is a more nuanced view of how resistance can arise: resistance is conferred if a mutation *increases* how well the natural substrate binds ($\Delta\Delta G_{\mathrm{DHA}} < 0$), as well as *decreasing* how well the antibiotic binds ($\Delta\Delta G_{\mathrm{TMP}} > 0$). It is likely, however, that large changes in the magnitude of $\Delta\Delta G_{\mathrm{DHA}}$ will affect the action and turnover rate of the enzyme and so, in practice, there will be a limit on how much a mutation can affect the binding of the natural substrate. Applying either of the above criteria generates a prediction of whether a mutation confers resistance or not and one of the aims of this paper is to assess if criterion R2 is more accurate and precise than R1. For either resistance criterion to classify a mutation as conferring resistance (or having no effect) the relevant free energy in R1 or R2 must be lie demonstrably one side of the 0.8 kcal/mol threshold

or the other; if the confidence limits bracket the threshold, then either criterion must return a classification of 'unknown'. This is a small departure from most culture-based microbiology tests which simply return a binary 'resistant' or 'susceptible' classification.

We chose a series of mutations in the chromosomal gene *dhfB* identified by whole-genome sequencing of *S. aureus* clinical infections from two hospitals in the UK (Gordon et al., 2014). As expected, by far the most common naturally occurring TMP resistance-conferring mutation in *S. aureus* DHFR was F99Y (Gordon et al., 2014; Dale et al., 1997). Several studies have shown that this common mutation reduces the binding free energy of TMP to *S. aureus* DHFR by 2.0 $\pm$ 0.2 kcal/mol (Dale et al., 1997; Oefner et al., 2009; Frey et al., 2010, 2012; Pires et al., 2015), equivalent to a 24 fold increase in the dissociation constant, K_{i}. This is a large effect given the mutation only replaces a hydrogen by a hydroxyl. Two further resistance-conferring mutations were chosen: L41F, which has also been previously observed (Vickers et al., 2009), and the double mutation F99Y/L21V, which has not – the related triple mutation F99Y/L21V/N60I was, however, identified as resistant 20 years ago (Dale et al., 1997). Mutating two residues simultaneously is likely to lead to convergence issues, and we therefore decomposed the double F99Y/L21V mutation into two separate mutations, F99Y and Y99L21V, summing the free energies to obtain the result for the double mutation (Klimovich et al., 2015). Although it has not yet been observed in isolation, we also calculated the effect of the isolated L21V mutation, allowing us to test the additivity of these mutations. Both the L41F and F99Y/L21V mutations are rare, only being observed once each among nearly 1,000 UK clinical isolates (Gordon et al., 2014). Any classification method must be able to distinguish true positives from true negatives, and therefore we also studied the effect of four mutations in *S. aureus* DHFR that were each detected multiple times in the isolate collection and had no effect on the action of TMP based on the results of conventional drug susceptibility testing. These were F123L, A135T, V76A and I83V (Fig. 1B) and are negative controls.

To confirm the phenotype of these seven mutations and to provide a consistent quantitative dataset, a subset of the clinical isolates that were sequenced as part of the previous study (Gordon et al., 2014) were retrieved and re-tested as described in the Methods. The TMP MICs were determined for each patient isolate (Table S1); up to five independent measurements were obtained, depending on how many clinical isolates of that mutation existed. The values obtained

agree well with both MIC values recorded by Public Health England during routine testing (Table S2) and those previously reported in the literature (Pires et al., 2015; Dale et al., 1997; Frey et al., 2010, 2012; Vickers et al., 2009).

Alchemical free energy calculations accurately predict which mutations confer resistance

Using our chosen alchemical free method (see Methods) we calculated how the free energy of binding of both TMP ($\Delta\Delta G_{\mathrm{TMP}}$) and DHA ($\Delta\Delta G_{\mathrm{DHA}}$) varies upon introducing each of the seven clinically-observed mutations. Thirty two values of $\Delta\Delta G_{\mathrm{TMP}}$ and $\Delta\Delta G_{\mathrm{DHA}}$ were calculated for each mutation, making 512 $\Delta\Delta G$ values in total. Since each pair of ($\Delta\Delta G_{\mathrm{TMP}}$,$\Delta\Delta G_{\mathrm{DHA}}$) values necessitated the calculation of 13 different ΔG values (Fig. S5), that makes 3,328 separate free energies. Since they originate from separate sets of simulations, each $\Delta\Delta G$ value is assumed to be independent, and therefore it is straightforward to examine how the values of $\Delta\Delta G_{\mathrm{TMP}}$ and $\Delta\Delta G_{\mathrm{DHA}}$ converge as the number of calculations, n, increases (Fig. 2). As expected, the uncertainty in the free energy is a maximum around $n = 3$ and then falls as the number of calculations is increased. The mutations with the largest confidence intervals are also those which perturb the largest number of atoms (F99Y/L21V, L41F and F123L).

The above analysis assumes that each $\Delta\Delta G$ calculation is itself converged; the standard way to test this would be to compare the forward and reverse cumulative averages of each $\Delta\Delta G$ value (Yang et al., 2004; Klimovich et al., 2015). This is not possible here due to the large numbers of $\Delta\Delta G$ values; instead we demonstrate that increasing or decreasing the proportion of each simulation that is discarded does not significantly alter either the calculated numerical values, or the resulting classification (Fig. S1 & S2).

Whilst our predicted value of $\Delta\Delta G_{\mathrm{TMP}}$ for the common F99Y mutation (1.5 ± 0.2 kcal/mol) Fig. 3A, Table S3) does not agree within error with the mean value (2.0 ± 0.2 kcal/mol) of several previously published isothermal titration calorimetry (ITC) measurements (Pires et al., 2015; Oefner et al., 2009; Dale et al., 1997; Frey et al., 2010, 2012), there is considerable overlap between the predicted and experimental values. Furthermore, all three known resistance-conferring mutations (F99Y, F99Y/L21V and L41F) are predicted to reduce how well TMP

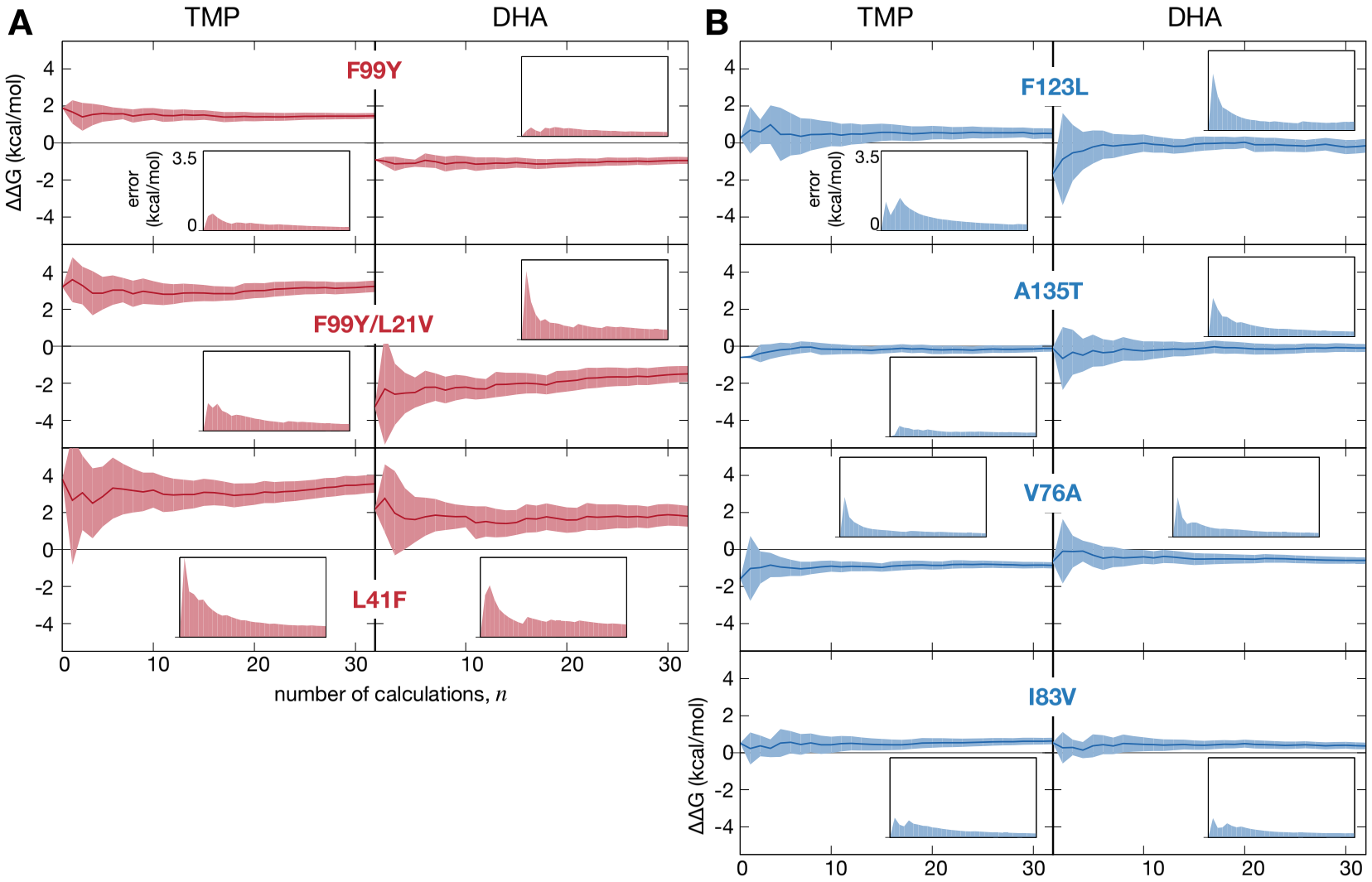

Figure 2: The calculated values for how the binding free energies change upon mutation ($\Delta\Delta G$) converge as the number of independent thermodynamic integration calculations is increased. Thirty-two separate calculations of $\Delta\Delta G_{\mathrm{TMP}}$ and $\Delta\Delta G_{\mathrm{DHA}}$ were run for each of the (*A*) three resistance-conferring and (*B*) four susceptible mutation (Table S3, S4). For each mutation, the variation in the mean $\Delta\Delta G$ value and its 95% confidence limits (calculated using the appropriate t-statistic) are shown as a function of the number of calculations, n. The inset graphs show how the confidence limits vary with n and all have the same scale. The initial 20% of each simulation has been discarded to avoid equilibration effects.

binds to DHFR ($\Delta\Delta G_{\mathrm{TMP}} > 0$, Fig. 3B, Table S3). Since these mutations were predicted to, on average, increase $\Delta\Delta G_{\mathrm{TMP}}$ by significantly more than 0.8 kcal/mol, they are classified as conferring resistance to TMP by criterion R1. Of the four negative control mutations, three are predicted to have 'no effect' on the action of TMP, although the 0.8 kcal/mol threshold is just outside the confidence limits for the F123L mutation. Since the 95% confidence limits for the remaining I83V mutation cross the threshold, this mutation is classified as having an 'unknown' phenotype.

But how do the mutations affect the binding of the natural substrate, DHA? In contrast to the binding of TMP, all the mutations, with the exception of L41F and I83V, are predicted to either have no effect on the binding of DHA, or to increase how strongly DHA binds to DHFR (Fig. 3C, Table S4). By considering the mean values for all four no-effect mutants, we find they are not predicted to change the magnitude of $\Delta\Delta G_{\mathrm{DHA}}$ by more than 0.5 kcal/mol, in line with our expectation that $\Delta\Delta G_{\mathrm{DHA}} \sim 0$.

Plotting the mean values of $\Delta\Delta G_{\mathrm{DHA}}$ against $\Delta\Delta G_{\mathrm{TMP}}$ (Fig. 4) allows us to classify the seven mutations using the second resistance criterion (R2). This condition predicts that all three known resistance-conferring mutations confer resistance to TMP, whilst of the four negative controls, three (V76A, A135T and I83V) are correctly predicted to have no effect on the action of TMP. Since the confidence limits of the remaining F123L mutation straddle the 0.8 kcal/mol threshold, it is predicted to have an 'unknown' effect. If the natural substrate binds more strongly to the enzyme ($\Delta\Delta G_{\mathrm{DHA}} < 0$), one could hypothesise that this should improve the turnover rate, if binding is the rate-limiting step. We speculate that L41F and I83V (especially the former) induce a fitness cost, since they reduce how well DHA binds to DHFR, whilst V76A, F99Y and particularly F99Y/L21V, bring a fitness benefit, with the others have no effect on the fitness of the enzyme. Since the free energies for the L21V and Y99L21V mutations (Table S3 & S4) are identical, to within error, we conclude that the effects of the F99Y and L21V mutations on the binding of TMP or DHA in the double F99Y/L21V mutant are additive

.

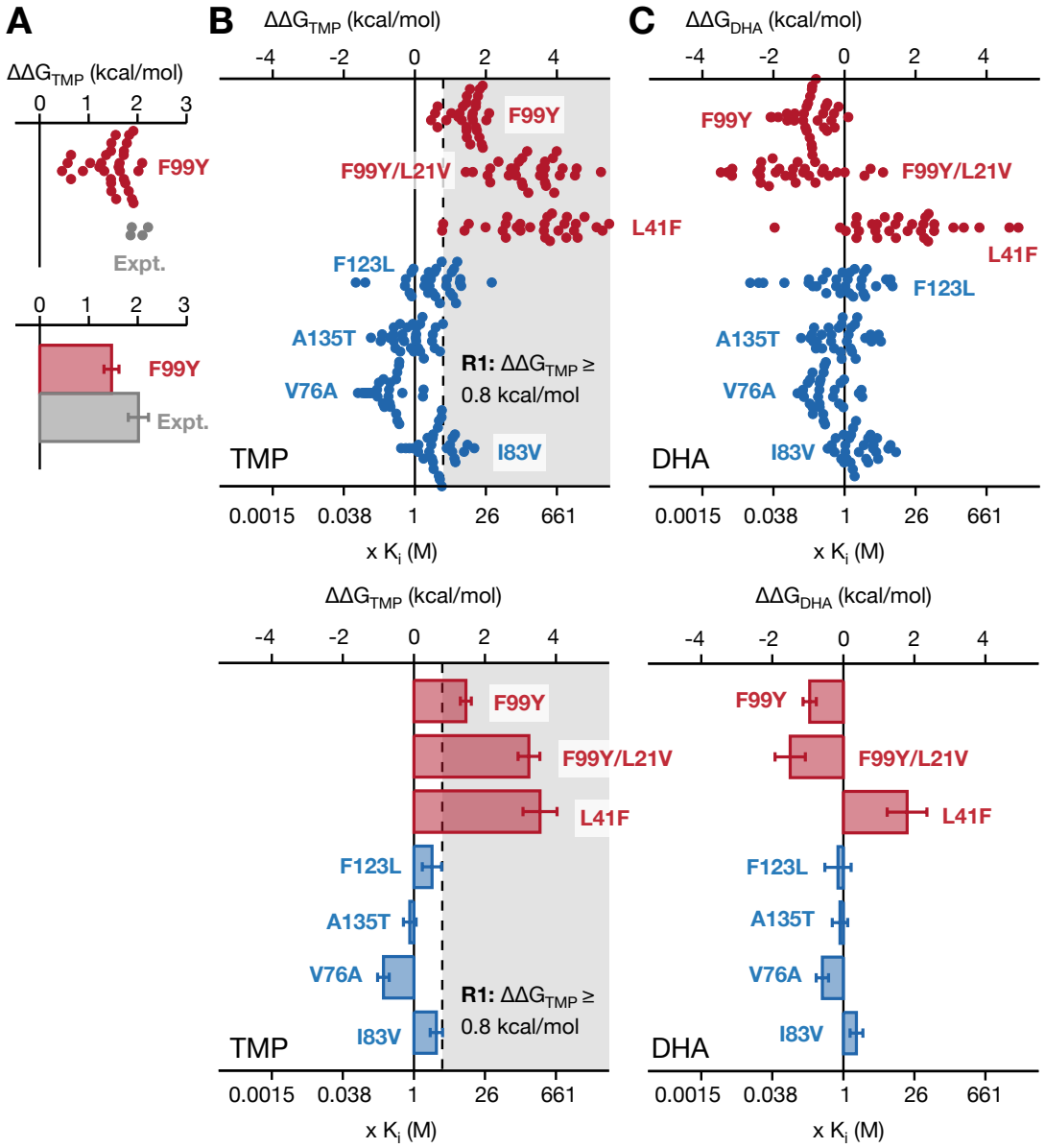

Figure 3: Thermodynamic integration correctly calculates how much the F99Y mutation reduces the TMP binding free energy and the R1 resistance criterion correctly classifies 6 of the 7 clinical mutations. (*A*) Whilst the predicted change in the binding free energy of TMP ($\Delta\Delta G_{\mathrm{TMP}}$) due to the F99Y mutation does not agree with previously published experimental data, the difference is small. (*B*) Applying resistance criterion R1 correctly classifies the F99Y, F99Y/L21V and L41F mutations as conferring resistance to TMP. The mutation L21V is also predicted to confer resistance. Of the four mutations known to have no effect on the action of TMP, F123L, A135T and V76A are correctly classified as not conferring resistance and I83V is classified as having an unknown effect. The fold change in the dissociation equilibrium constant (K_{i}) is also shown. Each value of $\Delta\Delta G$ is the mean of 32 independent calculations (Tables S3, S4), and the bars represent 95% confidence limits, using the appropriate t-statistic. The initial 20% of each simulation has been discarded to avoid equilibration effects. Discarding 10% or 50% of the data does not alter these conclusions (Fig. S1 & S2). (*C*) The same calculations were repeated, but with dihydrofolic acid (DHA) bound. With the exception of L41F and I83V, no mutation decreases how well DHA binds to DHFR, to within error.

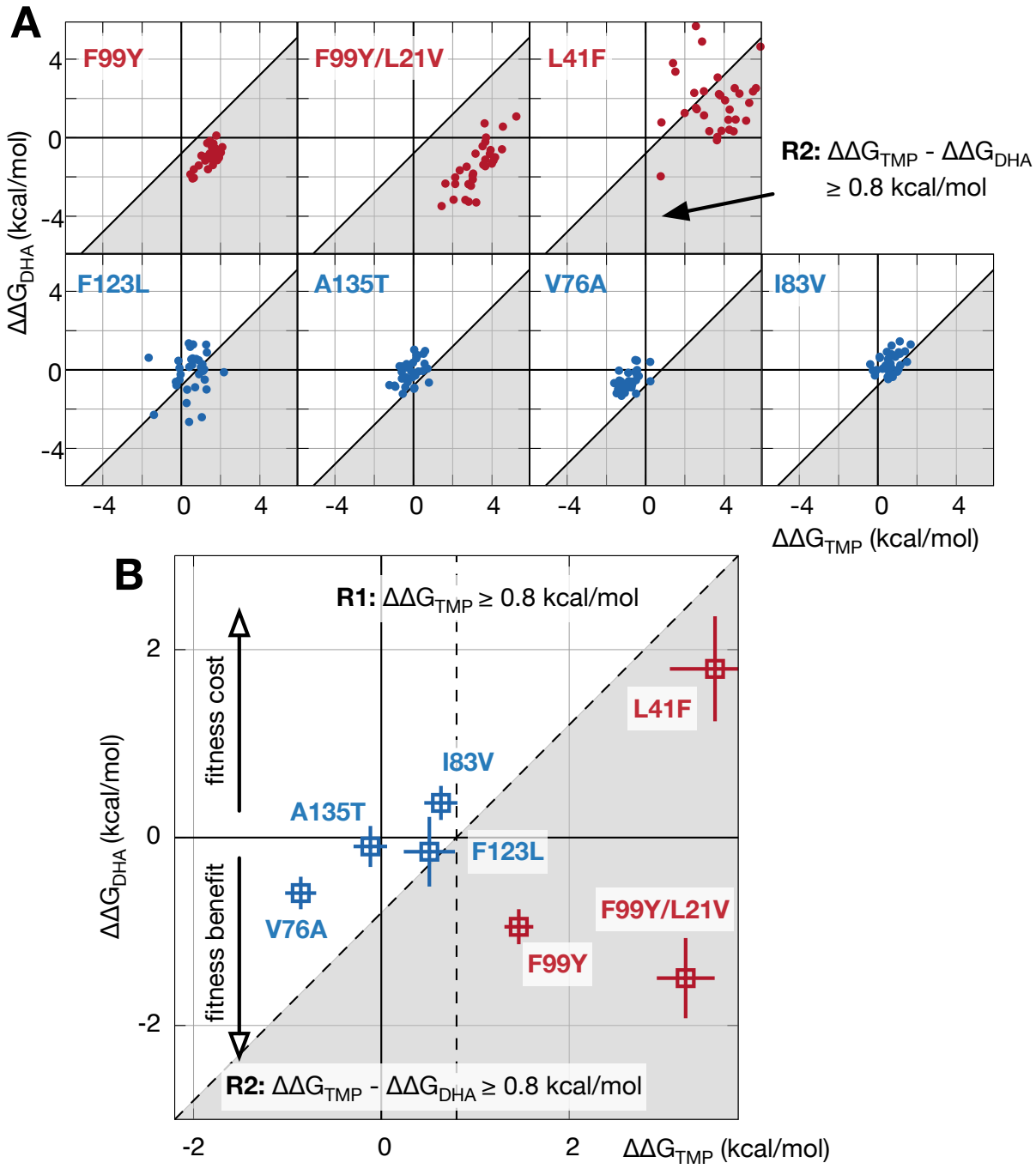

Figure 4: The R2 resistance criterion correctly predicts the effect of six of the seven mutations, with F123L being classified as having an unknown effect. (*A*) Plots of how each mutation is predicted to affect the binding of DHA against TMP (i.e. Fig. 3B v. C) for each of the 32 independent pairs of calculations. The region defined by the R2 resistance criterion is shaded grey. (*B*) Plotting the mean values with 95% confidence intervals demonstrates that the R2 resistance criterion correctly classifies all bar the F123 mutation which is predicted to have an unknown effect. The variation with n is shown in Fig. S3. All mutations are colored according to the same scheme as Fig. 1.

Predicting minimum inhibitory concentrations.

A stronger test of our approach is to compare against quantitative, rather than qualitative, data for *all* the mutations tested, rather than just F99Y. In the absence of quantitative binding data for the other mutations (as measured by e.g. ITC), we can instead predict the MIC for each mutation using Equation S2 and then compare it to the experimentally observed mean MICs (Table S1). As described in the Methods, the TMP MICs were measured by bioMérieux E-test. These have a roughly-doubling ladder of antibiotic concentrations going from 0.002 to 32 mg/L, a range of 16,000 fold. At first glance, there is a good correlation between the predicted and observed MICs (Fig. 5). This is, however, not a thorough test since (i) the experimental values have an upper limit of > 32 mg/ml and so we cannot distinguish between the different resistance-conferring mutations and (ii) there are no mutations that confer an intermediate level of resistance. Despite this, five of the seven predicted MICs can be said to be in 'essential agreement', since they are within a single doubling dilution (within the $2\times$ lines) of the reference method value (ISO, 2007) and, overall, it is promising that it appears possible to predict MICs to within a factor of 2-4.

We conclude that alchemical free energy methods are not only able to distinguish resistance-conferring mutations from susceptible mutations but also, by comparing to ITC data and MIC data, can make quantitatively accurate predictions, although more work is required before it will be possible to confirm that one can formally relate $\Delta\Delta G$ values to MICs. This proof of principle also study suggests that a good level of confidence in the phenotype of a mutation can be obtained by only predicting the effect on the binding of the antibiotic (i.e. criterion R1), in this case trimethoprim.

Classifying mutations using an alchemical free energy method is sensitive and specific.

Given predictions made by this type of approach could, one day, be used to drive clinical decision making, it is essential to establish the sensitivity and specificity of the method. First, let us assume that our sets of 32 pairs of $\Delta\Delta G_{TMP}$ and $\Delta\Delta G_{DHA}$ values per mutation are representative. The classification performance of the method can then be modelled by repeatedly drawing (with

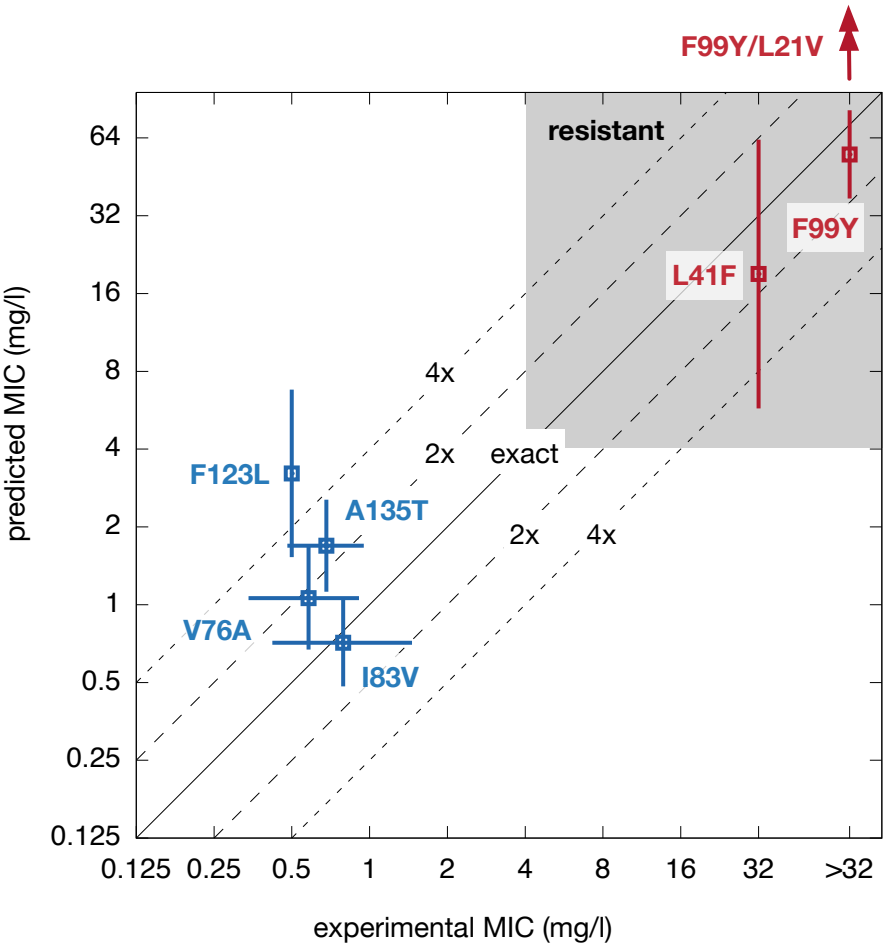

Figure 5: Our predicted values of the minimum inhibitory concentration (MIC) of trimethoprim correlate moderately well with the experimentally measured MICs. The predicted values are inferred from how the mutation alters the relative binding of trimethoprim (TMP) and dihydrofolic acid using Equation S4. Note that since the bioMérieux E-test does not measure above 32 mg/l, we have likewise cut off our predicted values at > 32 mg/l. The predicted mean MIC for the F99YL21V mutation is very large ($\sim$ 2,400 mg/l) and therefore cannot be plotted in this range. Clinically, a *S. aureus* DHFR mutation is classified as resistant if the TMP MIC ≥ 4 mg/l (EUCAST, 2016); this region is shaded light grey. To aid interpretation, lines corresponding to a perfect correlation, and within factors of $2\times$ and $4\times$ are drawn. Since our calculations only yield a fold increase in the MIC, all the predicted MIC values are assumed to be relative to a wildtype (geometric mean) MIC of 1.1 mg/l (EUCAST, 2016). The mutations are colored according to the same scheme as Fig. 1.

replacement) samples containing n values of $\Delta\Delta G_{TMP}$ and n values of $\Delta\Delta G_{DHA}$ and applying either resistance criterion to produce a classification. We repeated this bootstrapping approach 10,000 times at each value of $2 \leq n \leq 32$ and a summary of the results at five distinct values of $n \in \{3, 5, 10, 16, 32\}$ is shown in Fig. 6 (see also Fig. S4). Interestingly, even at small values of n, the method is unlikely to return an incorrect categorical prediction – the highest false categorisation rate occurs when applying the R1 resistance criterion to the I83V mutation at $n = 3$, and even then our analysis suggests the method would have incorrectly classified this mutation as conferring resistance only 2.5 % of the time with an 'unknown' result being returned in 91% of cases. We conclude that the method is robust in the sense that once n is large enough for it to return a definite categorisation it is highly likely to be correct.

The performance of a binary classification process is usually assessed by considering the true positive and true negative rates of detection, often referred to as the sensitivity and specificity, respectively. These are given in Table 1. Since our approach gives a ternary classification ('unknown' in addition to 'resistant' and 'susceptible'), there are two ways one can define the sensitivity and specificity. The difference rises from whether one includes the uncharacterised cases in the numbers of false positives and false negatives, or whether these cases can be excluded, since the method has (correctly) not attempted a definitive classification. If we first consider the former, more conservative definition, then the sensitivities / specificities are relatively low at small values of n and increase with n, achieving 99.7 / 61.3% for the R1 resistance criterion and 78.6 / 72.8% for the R2 criterion at $n = 10$ before reaching 100.0 / 77.9% and 84.0 / 91.0% at $n = 32$, respectively (Table 1). However, the proportion of uncharacterised cases fall dramatically from 37% (55%) for the first (second) resistance criterion at $n = 3$, to 11% (13%) at $n = 32$. If all these cases are excluded then all the sensitivities and specificities are $\geq 98\%$, suggesting that (i) the increase in the conservative estimates of the sensitivities and specificities is entirely driven by the decrease in the proportion of uncharacterised cases and (ii) our previous observation that the method rarely incorrectly classifies a mutation is correct. We conclude that the main effect of increasing the number of free energy calculations used in a prediction is increasing the likelihood that a definite classification will be made. We cannot, though, conclude which resistance criterion is 'better' since both the R1 and R2 resistance criteria struggle to classify two mutations each (F123L & I83V and F123L & L41F, respectively), even at high

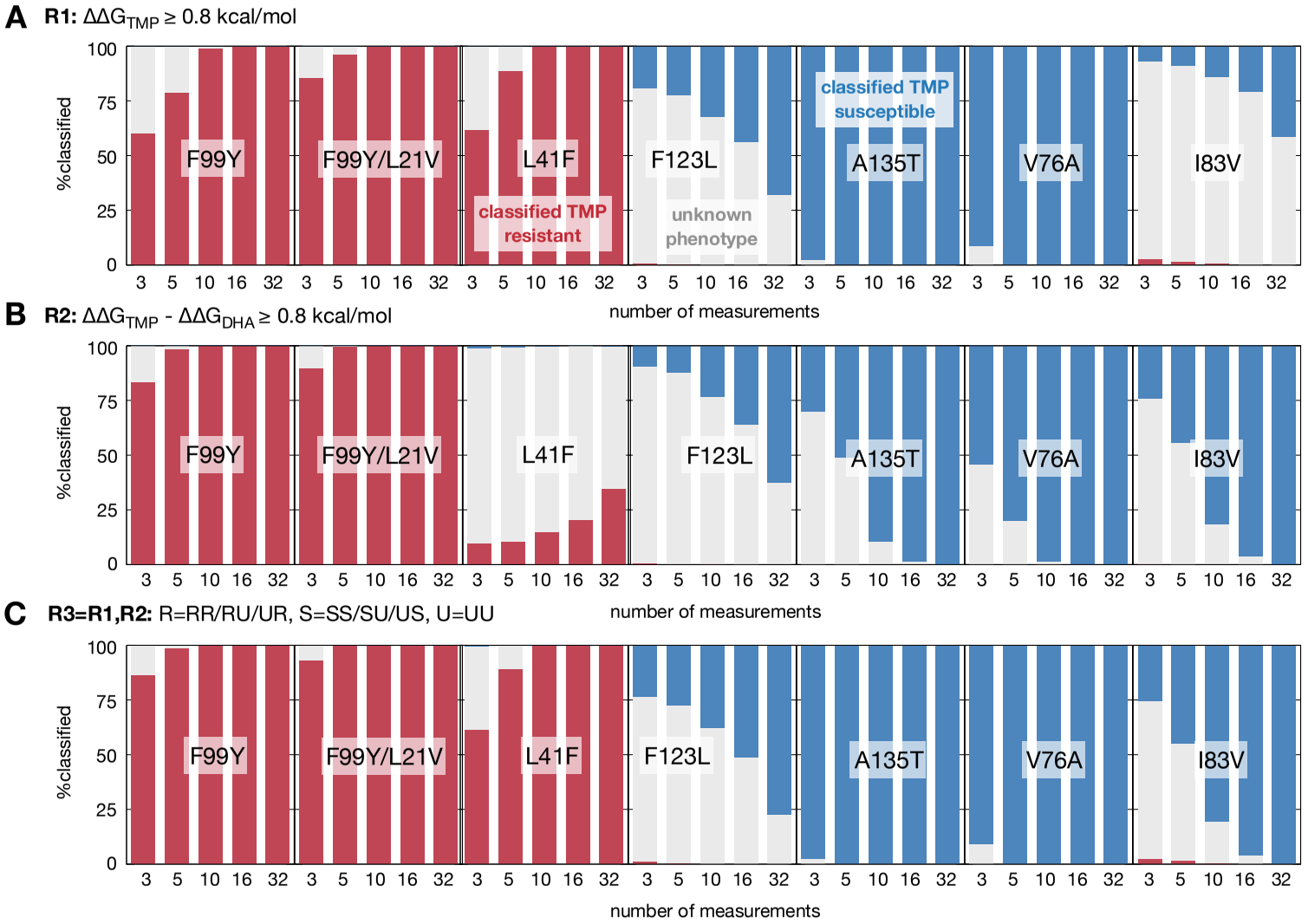

Figure 6: Predicting whether a mutation confers resistance is accurate and robust. The data in this figure were calculated by sampling-with-replacement and classifying 10,000 sets of n values of $\Delta\Delta G_{TMP}$ and n values of $\Delta\Delta G_{DHA}$ for $n \in \{3,5,10,16,32\}$ A classification is then made from each bootstrapped sample of free energies, and the results shown here as a function of n, the number of measurements in each sample, depending on whether the (A) first or (B) second resistance criterion was applied. (C) The results of applying both criteria and taking a consensus is also shown. See Fig. S4. How well the R1 & R2 criteria classify each mutation can be understood by considering the location and relative variations of each mutation on the $\Delta\Delta G_{DHA}$ versus $\Delta\Delta G_{TMP}$ plane. This is shown in Fig. 4 for $n = 32$ and examples of how it varies with n are shown in Fig. S3.

values of n. Difficulties in classifying a mutation are due to a combination of where it happens to fall relative to the two free energy thresholds on the $(\Delta\Delta G_{TMP}, \Delta\Delta G_{DHA})$ plane (Fig. 4, S3) and the variability between individual free energy calculations, which is related to the magnitude of the perturbation. The performance of either criteria therefore critically depends on which mutations have been selected to make up a test-set and, since we have only studied seven mutations, we cannot yet conclude which is preferable. Allowing a mutation to affect the binding of the natural substrate as well as the antibiotic is more elegant and hence one would expect the R2 resistance criterion to be more accurate, but it also requires $\frac{13}{8}\times$ the number of free energy calculations (Fig. S5).

One final possibility is to use the classifications from *both* criteria to make an ensemble prediction. The cases where both resistance criteria agree are trivial; the key question is how to classify mixed classifications e.g. RU. Here we assume that a definitive classification ('resistant' or 'susceptible') will overrule any 'unknown' classification and 'resistant' will overrule 'susceptible'. Hence if the results of applying the R1 and R2 resistance criteria can be represented as two letters, we shall define our ensemble rules for predicting resistance, susceptibility or unknown phenotypes as [RR,RU,UR,RS,SR], [SS,SU,US] and [UU], respectively. This ensemble method improves the classification performance, as measured by sensitivities and specificities (Table 1), for these seven mutations at least. It is, however, slightly unsatisfying since it weakens the link between the effect of the mutation on how well the antibiotic binds to the protein and the effectiveness of the drug.

(a) Resistance criterion R1. $\Delta\Delta G_{\mathrm{TMP}} \geq 0.8$ kcal/mol

	all		excluding uncharacterised		
n	sensitivity	specificity	sensitivity	specificity	uncharacterised
3	70.3 %	53.8 %	100.0 %	98.5 %	37.5 %
5	90.8 %	57.9 %	100.0 %	99.1 %	25.4 %
10	99.7 %	61.3 %	100.0 %	99.7 %	19.4 %
16	100.0 %	66.6 %	100.0 %	99.9 %	16.7 %
32	100.0 %	77.9 %	100.0 %	100.0 %	11.1 %

(b) Resistance criterion R2. $\Delta\Delta G_{\mathrm{TMP}} + \Delta\Delta G_{\mathrm{DHA}} \geq 0.8$ kcal/mol

	all		excluding uncharacterised		
n	sensitivity	specificity	sensitivity	specificity	uncharacterised
3	60.4 %	29.4 %	99.6 %	99.2 %	54.9 %
5	75.1 %	46.7 %	99.8 %	99.9 %	39.1 %
10	78.6 %	72.8 %	99.9 %	100.0 %	24.2 %
16	80.3 %	82.7 %	100.0 %	100.0 %	18.5 %
32	84.0 %	91.0 %	100.0 %	100.0 %	12.5 %

(c) Consensus. Taking (R1,R2) then: R=(RR,RU,UR), S=(SS,SU,US), U=UU

	all		excluding uncharacterised		
n	sensitivity	specificity	sensitivity	specificity	uncharacterised
3	82.3 %	59.3 %	99.8 %	98.4 %	28.6 %
5	96.8 %	68.0 %	99.9 %	99.2 %	17.3 %
10	100.0 %	79.5 %	100.0 %	99.8 %	10.2 %
16	100.0 %	86.8 %	100.0 %	99.9 %	6.6 %
32	100.0 %	94.3 %	100.0 %	100.0 %	2.8 %

Table 1: The expected proportion of classifications which would be returned with an 'unknown' phenotype decreases as the number of calculations, n, increases. The resulting sensitivities and specificities also increase with n. Two sets are given; the latter excludes all classifications with an unknown phenotype. All sensitivities and specificities are estimated by creating 10,000 samples of n values of $\Delta\Delta G_{\mathrm{TMP}}$ and n values of $\Delta\Delta G_{\mathrm{DHA}}$ by drawing-with-replacement from the larger set of 32 calculations. Results are given for the (a) R1 and (b) R2 resistance criteria. (c) Applying a consensus where any definitive 'resistance' or 'susceptible' classification overrules any 'unknown' classification is optimal.

DISCUSSION

We have shown that alchemical free energy methods can predict whether mutations in *S. aureus* DHFR confer resistance or not to the antibiotic trimethoprim. This paves the way for the introduction of such structural-based *predictive* methods into a genetics-based clinical microbiology service (Didelot et al., 2012; Köser et al., 2014) – allowing novel or insufficiently-characterised mutations to be assessed, thereby mitigating one of the key weaknesses of genetics-based clinical microbiology. The potential benefits of transitioning from laboratory- to genetics-based microbiology in the clinical setting are large: a reduction in the time for drug susceptibility testing (especially for slow-growing pathogens such as *Mycobacterium tuberculosis*), automatic epidemiological monitoring of the dispersal of specific resistance mechanisms and ever-decreasing cost. The switch to a genetics-based clinical microbiology will ultimately lead to increased precision in antibiotic prescribing and reduced selection for antibiotic resistance. The clinical transition has just begun: in early 2017 Public Health England adopted whole-genome sequencing for routine drug susceptibility testing for *M. tuberculosis* infections (Walker et al., 2017) and other countries look likely to follow suit.

Establishing the accuracy and reproducibility of any predictive method is essential, especially if it could ultimately drive decisions in a clinical setting. We emphasise the vital importance of (i) having negative controls, which here was enabled by a previous clinical whole-genome sequencing (WGS) study (Gordon et al., 2014), (ii) running multiple repeats, which has the additional benefit of simplifying the estimation of errors (Coveney and Wan, 2016), and (iii) systematically assessing the sensitivity and specificity of any method.

Ultimately, for predictions made by a computational method such as ours to form part of an antimicrobial diagnostic workflow, it must satisfy the same standards as any new lab-based diagnostic method (ISO, 2007; U.S. Department of Health and Human Services Food and Drug Administration, 2009). The key metrics used to assess a new method are the major discrepancy (MD) rate (the proportion of cases where the reference method predicts the infection is sensitive to an antibiotic but the new method predicts it is resistant) and the very major discrepancy (VMD) rate (which is the proportion of cases the reference method predicts the infection is resistant but the new method predicts it is sensitive). For a diagnostic test to be approved by

the International Standards Organization, both the MD and VMD $< 3\%$. As noted earlier, our method very rarely produces an incorrect definitive classification, and hence if 'unknown' results can be excluded, our method, based on the results in this paper, satisfies these criteria. For example, if we take a 'worst' case and consider only $n = 3$ then the VMD and MD for the first resistance criterion are 0.0 % & 1.6 %, respectively, whilst for the second resistance criterion the VMD and MD are 0.4 % & 0.8%. In making this comparison, we are not claiming that this method is sufficiently accurate for use in a clinical microbiology workflow for diagnosing antibiotic resistant infections – clearly many more mutations and proteins need to be tested – but rather, in combination with the sensitivity and specificity analysis, it does show that this method has the potential to predict the effect of novel and rare mutations on the action of antibiotics.

That the very major discrepancy rate is generally low but the proportion of classifications that are returned with an 'unknown' phenotype falls as n, the number of free energy calculations used to make a prediction, increases, suggests that a sensible way of applying this method would be to initially run a small number of free energy calculations (say $n = 5$) and try classifying the effect of the mutation. If a definitive result is returned, our analysis suggests that it is probably correct and will not be altered by adding more data. Alternatively, if the method cannot classify the effect of the mutation, then one can run additional free energy calculations until a definitive 'resistant' or 'no effect' classification can be made. In this way, some mutations would be classified very quickly, and others, like F123L or L41F, would take longer, as one would expect given the larger number of atoms being perturbed by the protein mutation.

Our approach has several weaknesses. Firstly, it assumes we know at a molecular level how an antibiotic works, specifically that it is a competitive inhibitor of an essential gene and it is mutations in that gene that we wish to examine; this is often, but not always, true. Secondly, it requires a high resolution experimental structure of the relevant bacterial protein with the antibiotic bound. Although the structural coverage of many bacterial genomes has more than doubled in the last ten years, with some species now having the structures of over half their proteins determined (Khafizov et al., 2014), the structural coverage of many pathogenic species remains low. In common with all applications of classical molecular dynamics, we are making two further key assumptions; (i) that our description of the molecular interactions is sufficiently accurate and (ii) that we have adequately sampled the phase space of the molecules. The first

is mitigated somewhat since it is protein atoms that are perturbed in the alchemical free energy calculation, and the protein forcefield has been extensively optimised (unlike in drug discovery where the atoms of a ligand, which inevitably are less well described, are perturbed). The second is mitigated by repeating calculations and allowing neighbouring simulations to exchange their Hamiltonians according to a Metropolis criterion. It is also difficult to calculate the relative free binding energy for some mutations using alchemical free energy methods; those perturbing large numbers of atoms are, as we have seen for e.g. F123L, take longer to converge. Finally, unlike in drug discovery where binding free energies (or equivalently dissociation equilibrium constants) are reported and to which one can directly compare predicted values of $\Delta\Delta G$, there is a paucity of binding free energy measurements for antibiotics. Instead the discipline of clinical microbiology measures and reports MIC values. It is possible, as we have done here, to relate the MIC to how the binding free energy changes upon the introduction of the mutation, but this requires several assumptions and is necessarily less direct.

Throughout this study we have calculated each component free energy (Equation S10 & Fig. S7) using the same number of λ simulations for the same duration, regardless of what type of free energy is being calculated and the size of the mutation being studied. This is almost certainly highly inefficient; in future work we will examine how to optimise our approach so that the minimum amount of computational resource is required to produce an accurate classification in the shortest time possible. This will include determining if a large number of relatively short simulations (as done here) is more accurate, at least when it comes to classifying, than a smaller number of longer simulations. Although some progress has been made in recent years examining this question in the context of endpoint free energy methods (Coveney and Wan, 2016), it has not yet been addressed for alchemical free energy calculations in general. Finally, it is only through the successful application of our approach to other proteins in other clinically-important pathogens where resistance is increasingly a problem, that it will be possible to determine if our method, or another one like it also based on the chemistry and structure of proteins, could, one day, be integrated into a genetics-based clinical microbiology pipeline.

Significance

The discovery of antibiotics was one of humanity's greatest achievements in the twentieth century; however, the evolution of antibiotic resistance by pathogens now threatens many advances of modern medicine. There is an urgent need for improved diagnostic tools so that resistant infections can be identified and treated appropriately. Analysis of whole-genome sequence data generated on affordable high-throughput platforms has the potential to allow resistant infections to be more rapidly and cheaply diagnosed in the clinic than conventional culture based approaches. A key limitation of this approach is that it cannot identify whether rare or previously unseen mutations will be associated with drug susceptibility or resistance. Since many antibiotics are competitive inhibitors, we hypothesise that mutations that confer resistance reduce how well the drug binds the target protein, whilst not significantly altering the binding free energy of the natural substrate. In this case, predicting whether a mutation confers resistance is equivalent to calculating the effect of the mutation on the binding free energies of both the antibiotic and the natural substrate. By relating these quantities to the standard clinical microbiology metric, the minimum inhibitory concentration (MIC), we are able to derive two different clinically-based criteria for classifying the effect of mutations and show that alchemical free energy methods, a well-established class of methods from computational chemistry, can not only predict which mutations confer resistance to trimethoprim, but are also quantitatively accurate.

Author contributions

PWF, NCG & ASW designed the study. KC & MJL tested the clinical isolates. AMK provided data from Public Health England. PWF setup, ran and analysed the simulations. PWF, MJL, TEAP, DWC & ASW wrote the paper.

Acknowledgement

The research was funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). We are grateful to the Science and Technology Facilities Research Council and Amazon Web Services for providing computer time. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

Abel, R., Mondal, S., Masse, C., Greenwood, J., Harriman, G., Ashwell, M.A., Bhat, S., Wester, R., Frye, L., Kapeller, R., et al. (2017). Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. Curr Opin Struct Biol *43*, 38–44.

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1-2*, 19–25.

Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R., and van Gunsteren, W.F. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem Phys Lett *222*, 529–539.

Blair, J.M.A., Webber, M.A., Baylay, A.J., Ogbolu, D.O., and Piddock, L.J.V. (2014). Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol *13*, 42–51.

Bradley, P., Gordon, N.C., Walker, T.M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L.J., Anson, L., de Cesare, M., Piazza, P, Votintseva, A.A., Golubchik, T., Wilson, D.J., Wyllie, D.H., Diel, R., Niemann, S., Feuerriegel, S., Kohl, T.A., Ismail, N., Omar, S.V., Smith,

E.G., Buck, D., McVean, G., Walker, A.S, Peto, T.E.A., Crook, D.W., and Iqbal, Z.(2015). Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nature Comms *6*, 10063.

Chodera, J.D., Mobley, D.L., Shirts, M.R., Dixon, R.W., Branson, K., and Pande, V.S. (2011). Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol *21*, 150–160.

Coveney, P.V. and Wan, S. (2016). On the calculation of equilibrium thermodynamic properties from molecular dynamics. Phys. Chem. Chem. Phys. *18*, 30236–30240.

Dale, G.E., Broger, C., D' Arcy, A., Hartman, P.G., DeHoogt, R., Jolidon, S., Kompis, I., Labhardt, A.M., Langen, H., Locher, H., Page, M.G.P., Stüber, D., Then, R.L., Wipf, B., Oefner, C. (1997). A single amino acid substitution in Staphylococcus aureus dihydrofolate reductase determines trimethoprim resistance. J Mol Biol *266*, 23–30.

Davies, S.C. (2013). Annual Report of the Chief Medical Officer - Vol 2. Technical report, Department of Health, UK Government.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E.A., and Crook, D.W. (2012). Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genetics *13*, 601–12.

Dotson, D.L., Seyler, S.L., Linke, M., Gowers, R.J., and Beckstein, O. (2016). datreant: persistent, Pythonic trees for heterogeneous data. In S. Benthall and S. Rostrup, eds., Proc 15th Python Sci Conf, 51–56.

EUCAST (2016). Data from the European Committee on Antimicrobial Susceptibility MIC distribution website. http://www.eucast.org.

Fowler, P.W., Geroult, S., Jha, S., Waksman, G., and Coveney, P.V. (2007). Rapid, accurate, and precise calculation of relative binding affinities for the SH2 domain using a computational grid. J Chem Theory Comput *3*, 1193–1202.

Fowler, P.W., Jha, S., and Coveney, P.V. (2005). Grid-based steered thermodynamic integration accelerates the calculation of binding free energies. Phil Trans R Soc Lond A *363*, 1999–2015.

Frey, K.M., Lombardo, M.N., Wright, D.L., and Anderson, A.C. (2010). Towards the understanding of resistance mechanisms in clinically isolated trimethoprim-resistant, methicillin-resistant Staphylococcus aureus dihydrofolate reductase. J Struct Biol *170*, 93–97.

Frey, K.M., Viswanathan, K., Wright, D.L., and Anderson, A.C. (2012). Prospective screening of novel antibacterial inhibitors of dihydrofolate reductase for mutational resistance. Antimicrob Agent Chemo *56*, 3556–3562.

Gapsys, V., Michielssens, S., Peters, J.H., Groot, B.L.D., and Leonov, H. (2015a). Calculation of Binding Free Energies. In A. Kukol, ed., Molecular Modeling of Proteins, volume 1215 of *Methods in Molecular Biology*, chapter 9, 173–209 (New York, NY: Springer New York).

Gapsys, V., Michielssens, S., Seeliger, D., and de Groot, B.L. (2015b). pmx: Automated protein structure and topology generation for alchemical perturbations. J Comp Chem *36*, 348–54.

Gapsys, V., Michielssens, S., Seeliger, D., and de Groot, B.L. (2016). Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angew Chem Int Ed *128*, 7490–7494.

Gilson, M.K. and Zhou, H.X. (2007). Calculation of protein-ligand binding affinities. Ann Rev Biophys *36*, 21–42.

Gordon, N.C., Price, J.R., Cole, K., Everitt, R., Morgan, M., Finney, J., Kearns, A.M., Pichon, B., Young, B., Wilson, D.J., Llewelyn, M.J., Paul, J., Peto, T.E.A., Crook, D.W., Walker, A.S., Golubchik, T. (2014). Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. J Clin Microbiol *52*, 1182–91.

Heaslet, H., Harris, M., Fahnoe, K., Sarver, R., Putz, H., Chang, J., Subramanyam, C., Barreiro, G., and Miller, J.R. (2009). Structural comparison of chromosomal and exogenous dihydrofolate reductase from Staphylococcus aureus in complex with the potent inhibitor trimethoprim. Proteins *76*, 706–717.

ISO (2007). ISO 20776-2: Clinical laboratory testing and in vitro diagnostic test systems. Technical report, International Standards Organization.

Jefferys, E., Sands, Z.A., Shi, J., Sansom, M.S.P., and Fowler, P.W. (2015). Alchembed: A computational method for incorporating multiple proteins into complex lipid geometries. J Chem Theo Comp *11*, 2743–2754.

Khafizov, K., Madrid-Aliste, C., Almo, S.C., and Fiser, A. (2014). Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. Proc Natl Acad Sci U S A *111*, 3733–8.

Klimovich, P.V., Shirts, M.R., and Mobley, D.L. (2015). Guidelines for the analysis of free energy calculations. J Comp Aided Mol Des *29*, 397–411.

Köser, C.U., Ellington, M.J., and Peacock, S.J. (2014). Whole-genome sequencing to control antimicrobial resistance. Trends in Genetics *30*, 401–407.

Lenselink, E.B., Louvel, J., Forti, A.F., van Veldhoven, J.P.D., de Vries, H., Mulder-Krieger, T., McRobb, F.M., Negri, A., Goose, J., Abel, R., van Vlijmen, H.W.T., Wang, L., Harder, E., Sherman, W., IJzerman, A.P., Beuming, T. (2016). Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. ACS Omega *1*, 293–304.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–8.

Lowy, F. (2003). Antimicrobial resistance: the example of Staphylococcus aureus. J Clin Invest *111*, 1265–1273.

Michel, J., Foloppe, N., and Essex, J.W. (2010). Rigorous free energy calculations in structure-based drug design. Mol Inf *29*, 570–578.

Nurjadi, D., Olalekan, A.O., Layer, F., Shittu, A.O., Alabi, A., Ghebremedhin, B., Schaumburg, F., Hofmann-Eifler, J., Van Genderen, P.J.J., Caumes, E., Fleck, R., Mockenhaupt, F.P., Hermann, M., Kern, W.V., Abdulla, S., Grobusch, M.P., Kremsner, P.G., Wolz, C., Zanger, P. (2014). Emergence of trimethoprim resistance gene dfrG in Staphylococcus aureus causing human infection and colonization in sub-Saharan Africa and its import to Europe. J Antimicrobial Chem *69*, 2361–2368.

Oefner, C., Bandera, M., Haldimann, A., Laue, H., Schulz, H., Mukhija, S., Parisi, S., Weiss, L., Lociuro, S., and Dale, G.E. (2009). Increased hydrophobic interactions of iclaprim with Staphylococcus aureus dihydrofolate reductase are responsible for the increase in affinity and antibacterial activity. J Antimicrobial Chem *63*, 687–698.

Pankhurst, L.J., del Ojo Elias, C., Votintseva, A.A., Walker, T.M., Cole, K., Davies, J., Fermont, J.M., Gascoyne-Binzi, D.M., Kohl, T.A., Kong, C., Lemaitre, N., Niemann, S., Paul, J.. Rogers, T.R., Roycroft, E., Smith, E.G., Supply, P., Tang, P., Wilcox, M.H, Wordsworth, S., Wyllie, D., Xu, L., Crook, D.W. (2016). Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Resp Med *4*, 49–58.

Perez, A., Morrone, J.A., Simmerling, C., and Dill, K.A. (2016). Advances in free-energy-based simulations of protein folding and ligand binding. Curr Opin Struct Biol *36*, 25–31.

Pires, D.E.V., Blundell, T.L., and Ascher, D.B. (2015). Platinum: A database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. Nuc Acid Res *43*, D387–D391.

Price, N.C., Dwek, R.A., Ratcliffe, R.G., and Wormald, M.R. (2009). Principles and Problems in Physical Chemistry for Biochemists (Oxford University Press), third edition.

Samsudin, M.F., Parker, J.L., Sansom, M.S.P., Newstead, S., and Fowler, P.W. (2016). Accurate Prediction of Ligand Affinities for a Proton-Dependent Oligopeptide Transporter. Cell Chem Biol *23*, 299–309.

Shirts, M.R. and Chodera, J.D. (2008). Statistically optimal analysis of samples from multiple equilibrium states. J Chem Phys *129*, 124105.

Sugita, Y., Kitao, A., and Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations. J Chem Phys *113*, 6042–6051.

Tong, S.Y.C., Davis, J.S., Eichenberger, E., Holland, T.L., and Fowler, V.G. (2015). Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management. Clin Micro Rev *28*, 603–661.

U.S. Department of Health and Human Services Food and Drug Administration (2009). Guidance for Industry and FDA. Class II Special Controls Guidance Document : Antimicrobial Susceptibility Test Systems. Technical report.

Vickers, A.A., Potter, N.J., Fishwick, C.W.G., Chopra, I., and O'Neill, A.J. (2009). Analysis of mutational resistance to trimethoprim in Staphylococcus aureus by genetic and structural modelling techniques. J Antimicrobial Chem *63*, 1112–1117.

Walker, T.M., Cruz, A.L.G., Peto, T.E., Smith, E.G., Esmail, H., and Crook, D.W. (2017). Tuberculosis is changing. Lancet Infec Disease *17*, 359–361.

Walker, T.M., Kohl, T.A., Omar, S.V., Hedge, J., Del Ojo Elias, C., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K.E., Wilson, D.J., Clifton, D.A, Kapatai, G., Ip, C.L.C., Bowden, R., Drobniewski, F.A., Allix-Béguec, C., Gaudin, C., Parkhill, J., Diel, R., Supply, P., Crook, D.W., Smith, E.G., Walker, A.S., Ismail, N., Niemann, S., Peto, T.E.A., and Modernizing Medical Microbiology Informatics Group (2015). Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infec Disease *15*, 1193–202.

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J., Romero, D.L., Masse, C., Knight, J.L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D.L., Jorgensen, W.L., Berne, B.J., Friesner, R.A., and Abel, R. (2015). Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J Am Chem Soc *137*, 2695–2703.

Woods, C.J., Essex, J.W., and King, M.A. (2003). The Development of Replica-Exchange-Based Free-Energy Methods. J Phys Chem B *107*, 13703–13710.

World Economic Forum (2013). Insight Report: Global Risks, Eighth Edition.

Yang, W., Bitetti-Putzer, R., and Karplus, M. (2004). Free energy simulations: use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. J Chem Phys *120*, 2618–2628.

Zacharias, M., Straatsma, T.P., and McCammon, J.A. (1994). Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. J Chem Phys *100*, 9025.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Philip Fowler

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The clinical isolates tested in this study were collected and sequenced as described previously (Gordon et al., 2014).

METHOD DETAILS

Trimethoprim Susceptibility Testing

Susceptibility of test isolates to trimethoprim was determined by E-test (bioMérieux, Marcy l'Etoile, France) in accordance with the manufacturer's instructions. Breakpoints were interpreted according to EUCAST guidelines (EUCAST, 2016).

System building and equilibration

An experimental structure of *S. aureus* DHFR with trimethoprim (TMP) and NADPH bound (PDB:3FRE) was used to setup all simulations (Oefner et al., 2009). Apo structures were created by removing TMP. The generalized AMBER forcefield in conjunction with AMBER ff99SB-ILDN (Lindorff-Larsen et al., 2010) was used throughout and all simulations were carried out using GROMACS 5.0.x (Abraham et al., 2015). The mutations in the protein were represented using a dual topology and all GROMACS free energy topology files were prepared using `pmx` (Gapsys et al., 2015b). Each protein was solvated by adding waters and ions resulting in a simulation unit cell of dimensions 7.1 $\times$ 6.4 $\times$ 6.0 nm containing 27,077–27,120 atoms. For each mutant, separate apo, TMP- and DHA-bound short equilibration simulations were run. First the energy of each system was minimised using the steepest descent algorithm for 1000 steps, then the dynamics of the system evolved for 2.5 ns with an integration timestep of 1 fs. Electrostatic forces were calculated using the particle mesh Ewald method with a real space

cutoff of 1.2 nm. Van der Waals interactions were cutoff at 1.2 nm, with a switching function applied from 0.9 nm. A Langevin thermostat with a time constant of 2 ps was applied to keep the temperature at 310 K. The pressure was maintained at 1 atm by an isotropic Parinello-Rahman barostat with a time constant of 1 ps and a compressibility of 4.46×10^{-5} bar^{-1}. The lengths of all bonds involving a hydrogen were constrained using the LINCS algorithm. Since all the above simulations were run with $\lambda = 0$ (i.e. wildtype sidechain), we then ran a short simulation to 'phase-in' the mutant sidechain using the Alchembed procedure (Jefferys et al., 2015). This was repeated for different snapshots taken during the 2.5 ns equilibration trajectory and ensured that we had a range of starting conformations suitable for all the different alchemical and end-point simulations.

Alchemical simulations and calculations

A thermodynamic cycle was constructed (Fig. S7) and changes in the free energy of binding upon introduction of the mutation, $\Delta\Delta G$, was defined by a series of alchemical transformation free energies. We followed best practice and, when changing one sidechain into another, calculated three separate free energies (Klimovich et al., 2015). This was repeated first for the apo protein (ΔG_1) and then the complex (ΔG_6). First the electrical charges on the perturbing atoms are removed (ΔG_{11} & ΔG_{61}), before the van der Waals terms on the disappearing and appearing atoms are decoupled and coupled to the system, respectively (ΔG_{12} & ΔG_{62}), using a soft-core potential (Beutler et al., 1994; Zacharias et al., 1994). Finally the electrical charges on the new atoms are switched on (ΔG_{13} & ΔG_{63}). To keep the ligand within the active site, the distance between the protein and ligand centres of mass were restrained using a harmonic potential with a spring constant of 2000 kJ nm^{-1} mol^{-2}. The reference distances for TMP and DHA were 0.644 nm and 0.794 nm, respectively. The free energies of removing both restraints were calculated (ΔG_5 & ΔG_7). The final free energy is derived in the Supplemental Information and is given by

$$\Delta\Delta G = \Delta G_5 + (\Delta G_{61} + \Delta G_{62} + \Delta G_{63}) - (\Delta G_{11} + \Delta G_{12} + \Delta G_{13}) - \Delta G_7. \quad (1)$$

Each free energy was calculated by running either 8, 11 or 16 simulations at equally-spaced

values of the progress parameter, λ, between 0 and 1. To accelerate convergence, each set of 8, 11 or 16 simulations were coupled and attempted to exchange Hamiltonians every 1,000 steps (Sugita et al., 2000; Woods et al., 2003). Each set was run for 0.25 ns, meaning each free energy calculation required between 26 and 52 ns of molecular dynamics simulation. Thirty two pairs of $(\Delta\Delta G_{TMP}, \Delta\Delta G_{DHA})$ were calculated for each mutation (Table S3 and Fig. S5), 5 with $11\times\lambda$ values, 5 with $16\times\lambda$ values and 22 with $8\times\lambda$ values. No correlation between the number of λ values and the magnitude of the resulting value of $\Delta\Delta G$ was detected. Calculating 32 pairs of $\Delta\Delta G$ values for a single mutation therefore required 1.0 μs of molecular dynamics simulation. Eight mutations were calculated in total (since the F99YL21V mutation was decomposed into two separate mutations), making a total of 8.1 μs of molecular dynamics simulation. More daunting is that this is composed of 32,344 separate molecular dynamics simulations. These were stored and discovered using `datreant`, a flexible python module for handling heterogeneous file-based data (Dotson et al., 2016).

The first derivative of the internal energy at the specified value of λ, as well as the internal energy evaluated at all other values of λ were written to disc every 0.1 ps. This permitted the free energy (ΔG) to be calculated using either the multi-state Bennett acceptance ratio estimator (MBAR) (Shirts and Chodera, 2008) by the `alchemical-analysis` python module (Klimovich et al., 2015), or simple thermodynamic integration. Since no significant differences in $\Delta\Delta G$ values were observed, with the mean unsigned error in a value of $\Delta\Delta G$ being between 0.1-0.3 kcal/mol, depending on the number of atoms being perturbed, the latter was used for simplicity. A subset of the GROMACS input files is available for download allowing a single pair of $\Delta\Delta G$ values to be calculated for each mutant from https://github.com/philipwfowler/amr-free-energy-dhfr-examples.

The simulation parameters are the same as for the equilibration simulations above, except the tolerance factor for the Ewald sum is decreased to 10^{-6} to increase the accuracy of calculating electrostatic forces, as is standard in these types of calculations. To remove transient effects, the first 20% of each simulation was discarded. Discarding more (50%) or less (10%) of the data did not materially affect the results (Fig. S2).

QUANTIFICATION AND STATISTICAL ANALYSIS

Throughout, standard errors were calculated at a confidence level of 95%, taking into account the appropriate t-statistic for the sample size. This assumes each calculated value of ΔG is independent, which is reasonable since they are started from different initial structures taken from the equilibration simulations and run using different random seeds.

DATA AND SOFTWARE AVAILABILITY

The clinical isolates tested in this paper were sequenced in a previous study (Gordon et al., 2014) and, as a result, can be found in the European Nucleotide Archive Sequence Read Archive under study accession number ERP004655.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENITIFIER
Bacterial and Virus Strains		
Staphylococcus aureus: 20 clinical isolates	Gordon et al. (2014)	N/A
Staphylococcus aureus: 7 patient screening swab isolates	Gordon et al. (2014)	N/A
Critical Commercial Assays		
DNA extraction tissue kit S	DNA extraction tissue kit S	DT-S
Software and Algorithms		
GROMACS 5.0	Klimovich et al. (2015)	https://www.gromacs.org
AMBER ff99SB*ILDN forcefield	Lindorff-Larsen et al. (2010)	Distributed with GROMACS
pmx	Gapsys et al. (2015b)	https://github.com/dseeliger/pmx
datreant	Dotson et al. (2016)	https://github.com/datreant/datreant.core
alchemical-analysis	Klimovich et al. (2015)	https://github.com/MobleyLab/alchemical-analysis
Other		
Iso-Sensitest Agar (4mm depth)	Oxoid	PO0779
Etest Trimethoprim	BioMerieux	TR32US
HiSeq 2000 Sequencing System Gordon et al. (2014)	Illumina	HiSeq2000
Staphylococcus aureus DHFR protein structure	Protein Data Bank	3FRE

Supplemental Information

Mutation	Identifier	TMP ETest MIC
F99Y	C00003356	>32 mg/l
	C00007077	>32 mg/l
	C00007074	>32 mg/l
	C00013181	>32 mg/l
	C00000815	>32 mg/l
F99Y/L21V	C00000825	>32 mg/l
L41F	C00003323	32 mg/l
F123L	C00001107	0.5 mg/l
	C00003121	0.5 mg/l
	C00013402	0.5 mg/l
A135T	C00003374	0.38 mg/l
	C00001194	0.75 mg/l
	C00013217	0.5 mg/l
	C00008638	1 mg/l
	C00000857	0.38 mg/l
I83V	C00012815	0.75 mg/l
	C00008632	1 mg/l
	C00001110	0.75 mg/l
	C00001152	0.5 mg/l
	C00003290	0.5 mg/l
V76A	C00003280	0.38 mg/l
	C00013201	1 mg/l
	C00012748	1 mg/l
	C00012825	1 mg/l
	C00012751	1 mg/l

Table S1: Related to Figure 1. Clinical isolates used in this study and their trimethoprim minimum inhibitory concentrations (MIC), as measured by bioMérieux E-test. The F99Y/L21V and L41F mutations were only observed once, and hence each only has a single data point. Likewise the F123L mutation was only observed three times. For all other mutations five randomly-selected clinical isolates were tested as described in the Methods.

Mutation	mean MIC (mg/l)	MIC range (95% confidence)	Phenotype	$\Delta\Delta\Delta G_{\mathrm{TMP-DHA}}$ (kcal/mol)
F99Y	>32	–	Resistant	> 2.1
F99Y/L21V	>32	–	Resistant	> 2.1
L41F	32	–	Resistant	2.1
F123L	0.5	–	Susceptible	-0.5
A135T	0.6	0.3-0.9	Susceptible	-0.4 ± 0.3
V76A	0.9	0.4-1.5	Susceptible	-0.2 ± 0.4
I83V	0.7	0.5-1	Susceptible	-0.3 ± 0.2

Table S2: Related to Figure 1. Mean and confidence intervals for the measured trimethoprim minimum inhibitory concentrations (MIC), as measured by bioMérieux E-test, for the seven mutations chosen for this study. The mean MIC was calculated using the geometric mean and 95% confidence intervals are estimated using the appropriate t-statistic. The resulting phenotypes are consistent with the previously published study (Gordon et al., 2014).

Mutation	PHE MIC values data (mg/l)
F99Y	>32, >32
F99Y/L21V	–
L41F	–
F123L	0
A135T	0
V76A	0, 0, 0, 0, 1
I83V	0, 0, 1

Table S3: Related to Figure 1. The incidences and recorded trimethoprim minimum inhibitory concentrations (MIC) by routine monitoring by Public Health England (PHE). Neither the F99Y/L21V or L41F mutation were observed. All isolates containing any of the plasmid-encoded genes, *dfrA*, *dfrG* or *dfrK*, were excluded.

$\Delta\Delta G_{TMP}$	F99Y	Y99L21V	F99YL21V	L21V	L41F	F123L	A135T	V76A	I83V
01	1.88	1.31	3.20	2.44	3.81	0.27	-0.60	-1.60	0.52
02	1.47	2.53	4.00	1.77	1.52	1.11	-0.55	-0.45	-0.05
03	0.89	1.76	2.65	2.94	3.86	0.41	-0.03	-0.88	0.65
04	1.92	-0.28	1.63	0.94	0.80	2.17	0.02	-0.44	-0.15
05	1.82	1.03	2.85	2.78	4.29	-1.66	0.22	-1.36	1.68
06	1.46	2.43	3.89	2.05	5.62	0.51	0.04	-1.23	0.76
07	1.61	1.41	3.02	2.44	2.97	-0.28	0.48	-1.36	-0.26
08	0.64	0.80	1.44	3.52	2.60	1.07	0.04	-0.68	1.12
09	2.09	1.99	4.08	1.94	2.56	0.56	-0.96	-0.44	-0.38
10	1.91	0.21	2.12	1.87	4.05	0.76	-0.30	-0.52	0.41
11	0.57	1.48	2.05	1.67	0.77	-0.25	-0.21	-1.41	1.14
12	1.46	1.35	2.80	2.42	2.48	1.20	-0.35	-0.53	0.68
13	2.02	1.61	3.62	3.26	3.25	0.30	-0.35	-1.16	0.10
14	1.22	1.79	3.01	1.55	2.98	1.16	-0.59	-1.07	0.09
15	1.72	0.99	2.71	1.70	4.78	1.27	0.56	-1.48	0.24
16	0.64	1.71	2.35	2.62	2.87	0.58	0.15	0.24	0.36
17	0.46	2.47	2.93	1.46	1.99	-0.10	0.59	-0.35	0.46
18	1.63	1.18	2.81	2.02	1.40	-0.14	-0.92	-0.49	1.07
19	1.75	3.49	5.24	2.86	3.65	0.91	0.25	-0.71	1.49
20	1.03	2.63	3.66	1.77	3.63	1.30	-1.23	-1.34	0.70
21	1.56	1.60	3.16	2.20	5.29	0.39	-0.42	-0.74	0.76
22	1.26	3.30	4.56	2.06	2.56	-0.04	0.09	0.23	0.48
23	1.47	2.09	3.56	1.41	4.21	1.00	-0.41	-0.69	0.51
24	1.78	2.15	3.93	1.76	3.73	1.28	-0.71	-1.01	0.99
25	1.83	1.84	3.67	2.35	4.54	-0.09	0.71	-0.49	0.90
26	1.46	1.58	3.04	2.14	4.59	0.70	0.51	-0.90	1.13
27	1.55	1.97	3.52	1.76	5.47	0.45	-0.59	-1.27	0.75
28	1.36	0.77	2.12	2.45	4.26	0.60	-0.06	-1.05	0.50
29	1.64	2.06	3.69	1.89	5.85	1.04	0.15	-1.42	1.39
30	1.32	2.34	3.66	3.16	3.67	-1.40	0.04	-0.72	0.68
31	1.72	2.79	4.51	1.76	5.12	0.89	0.79	-0.86	0.53
32	1.72	2.44	4.16	2.16	4.49	0.38	-0.11	-1.26	1.05
	1.46	1.78	3.24	2.16	3.55	0.51	-0.12	-0.86	0.63
n=32	±	±	±	±	±	±	±	±	±
	0.15	0.29	0.31	0.21	0.48	0.27	0.18	0.17	0.18

Table S4: Related to Figure 2. The values of $\Delta\Delta G_{TMP}$ for the seven DHFR mutations (kcal/mol). Thirty-two thermodynamic integration calculations were run for each mutation. Each value is composed of eight separate free energies as per Equation 1 and the thermodynamic cycle (Fig. S5). The means and 95% confidence intervals are shown for both the first ten values and all thirty-two. The confidence interval takes into account the relevant t-statistic for the sample size.

$\Delta\Delta G_{DHA}$	F99Y	Y99L21V	F99YL21V	L21V	L41F	F123L	A135T	V76A	I83V
01	-0.91	-2.39	-3.30	-2.27	2.15	-1.69	-0.11	-0.67	0.55
02	-1.07	-0.25	-1.32	-0.10	3.36	-0.06	-1.22	0.47	-0.01
03	-1.41	-1.76	-3.17	-1.02	0.35	0.15	0.36	-0.15	0.32
04	-1.03	-1.32	-2.34	-0.42	0.78	-0.12	-0.96	-0.01	-0.28
05	-1.07	-1.31	-2.38	0.02	1.44	0.62	0.74	-1.14	1.30
06	-0.17	-0.66	-0.82	1.88	2.53	0.53	-0.92	-1.32	0.73
07	-1.40	-0.73	-2.13	-0.45	2.37	-0.61	-0.05	-0.03	-0.03
08	-2.05	-1.44	-3.48	0.20	1.45	0.28	1.03	-0.74	1.45
09	-0.48	-0.68	-1.16	-0.56	1.52	0.25	-0.80	-0.28	0.28
10	-0.98	-1.05	-2.02	-0.78	1.91	0.55	-0.63	-0.11	0.23
11	-2.08	-1.08	-3.16	0.54	-1.97	-0.79	0.21	-0.98	0.24
12	-0.81	-1.56	-2.37	-0.68	2.28	-0.50	0.12	0.50	0.14
13	-0.76	1.49	0.73	0.26	0.34	-1.01	0.29	-0.90	0.61
14	-1.17	-0.77	-1.94	0.51	1.14	0.03	-0.45	-1.17	0.67
15	-0.53	-0.95	-1.48	-1.45	2.24	1.28	0.31	-1.19	0.06
16	-1.62	-0.04	-1.66	0.50	4.90	0.58	0.59	-0.59	0.17
17	-1.86	-0.58	-2.45	0.10	1.26	0.09	0.96	-0.32	0.92
18	-0.91	-2.36	-3.27	0.31	3.80	0.46	-0.86	-0.53	0.89
19	-0.58	1.66	1.09	0.81	0.04	-0.22	-0.08	-0.60	0.41
20	-0.92	-0.20	-1.12	-1.94	-0.13	0.89	-0.78	-0.55	1.24
21	-0.49	-0.33	-0.82	1.41	1.78	-2.66	-0.88	-0.71	-0.36
22	-0.97	1.53	0.56	-0.17	5.69	-0.23	-0.29	0.41	0.02
23	-0.92	-0.45	-1.37	0.00	0.92	0.14	-0.37	-0.70	0.18
24	0.11	-0.74	-0.63	-0.36	2.22	-1.00	0.18	-0.70	-0.09
25	-1.14	-0.30	-1.45	-1.61	2.52	-0.51	0.07	-1.21	0.67
26	-0.91	-0.92	-1.83	-2.63	0.92	-0.89	0.82	-0.89	0.82
27	-0.37	-0.07	-0.44	-0.26	2.35	1.17	0.17	-1.03	-0.25
28	-1.61	-0.75	-2.36	3.97	0.41	1.29	-0.38	-0.97	-0.36
29	-0.30	0.30	0.01	-5.13	4.64	-2.42	0.76	-0.70	0.95
30	-0.28	0.06	-0.21	-2.60	3.07	-2.29	-0.16	-0.90	0.81
31	-0.60	0.01	-0.59	-3.93	0.87	0.49	-0.64	-0.57	-0.47
32	-1.10	0.10	-1.00	-3.75	0.32	1.35	-0.08	-0.59	0.02
	-0.95	-0.55	-1.50	-0.61	1.80	-0.15	-0.10	-0.59	0.37
n=32	±	±	±	±	±	±	±	±	±
	0.19	0.34	0.43	0.63	0.56	0.37	0.22	0.17	0.18

Table S5: Related to Figure 2. The values of $\Delta\Delta G_{DHA}$ for the seven DHFR mutations (kcal/mol). Thirty-two thermodynamic integration calculations were run for each mutation. Each value is composed of eight separate free energies as per Equation 1 and the thermodynamic cycle (Fig. S5). The means and 95% confidence intervals are shown for both the first ten values and all thirty-two. The confidence interval takes into account the relevant t-statistic for the sample size.

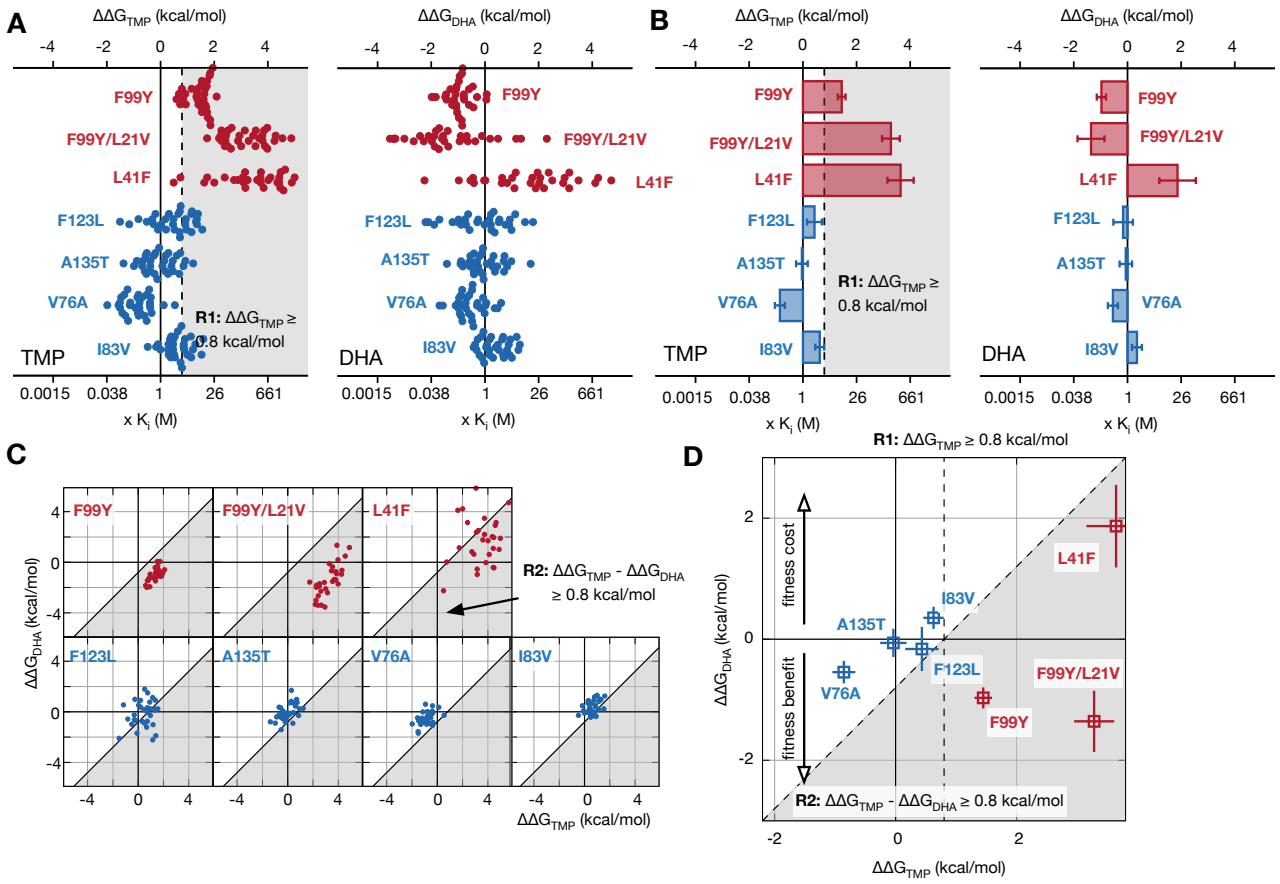

Figure S1: Related to Figures 3 & 4. Discarding less (10% compared to 25%) data from the alchemical simulations does not alter the classification of mutants. For a mutation to be classified as conferring resistance according either resistance criterion, the predicted change in binding free energy must lie in the shaded area. All values here are the mean of 32 independent simulations with 95% confidence limits calculated taking into account the appropriate t-statistic. Mutations are colored according to the same scheme as Figure 1.

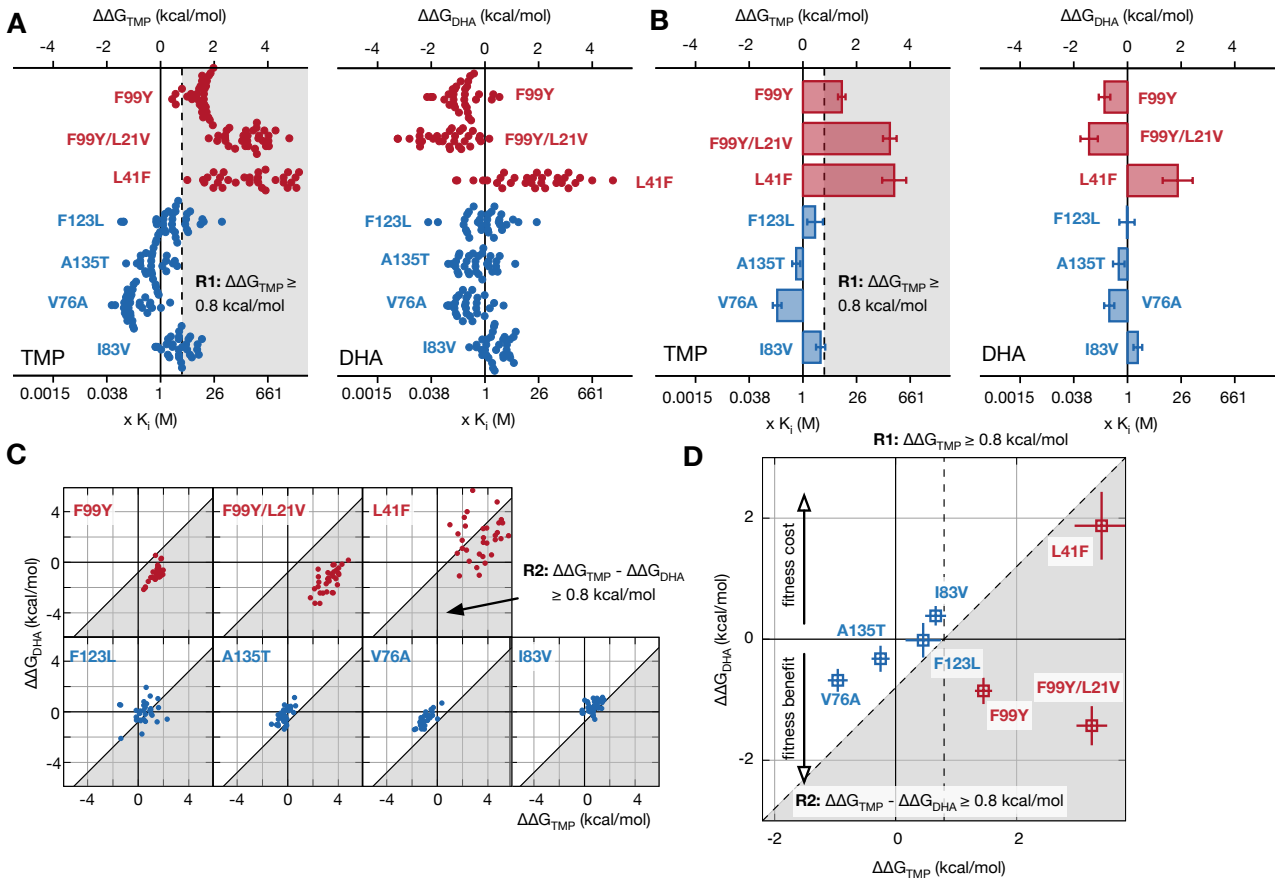

Figure S2: Related to Figures 3 & 4. Discarding more (50% compared to 25%) data from the alchemical simulations does not alter the classification of mutants. For a mutation to be classified as conferring resistance according either resistance criterion, the predicted change in binding free energy must lie in the shaded area. All values here are the mean of 32 independent simulations with 95% confidence limits calculated taking into account the appropriate t-statistic. Mutations are colored according to the same scheme as Figure 1.

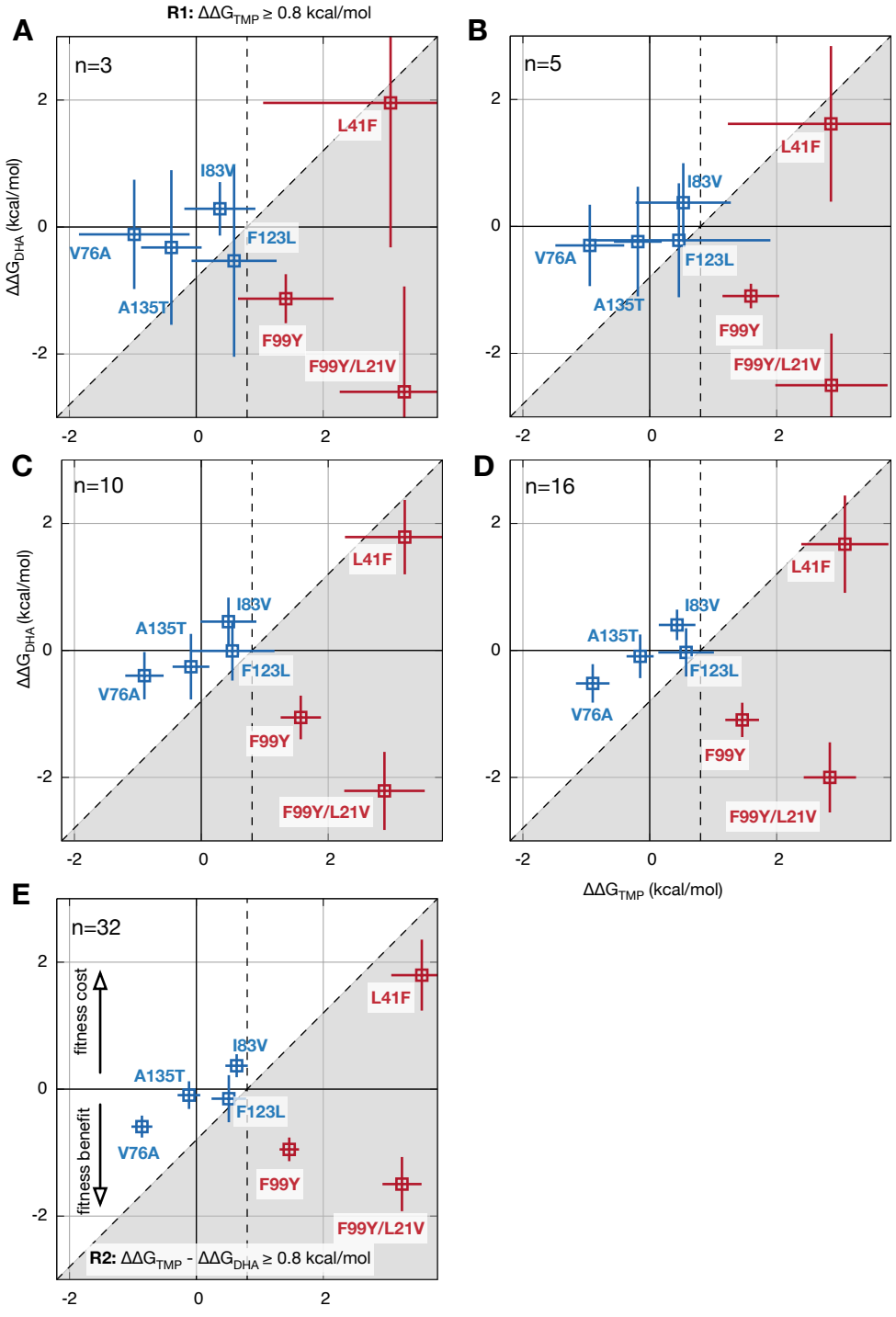

Figure S3: Related to Figure 4. Increasing the number of independent calculations better resolves where a mutation lies in the $\Delta\Delta G_{DHA}$ v. $\Delta\Delta G_{TMP}$ plane. As the number of calculations is increased from (A) $n = 3$, to (B) 5, (C) 10, (D) 16 and finally (E) $n = 32$ the 95% confidence limits reduce which decreases the chance that a mutation is either incorrectly classified, or classified as having an 'unknown' phenotype.

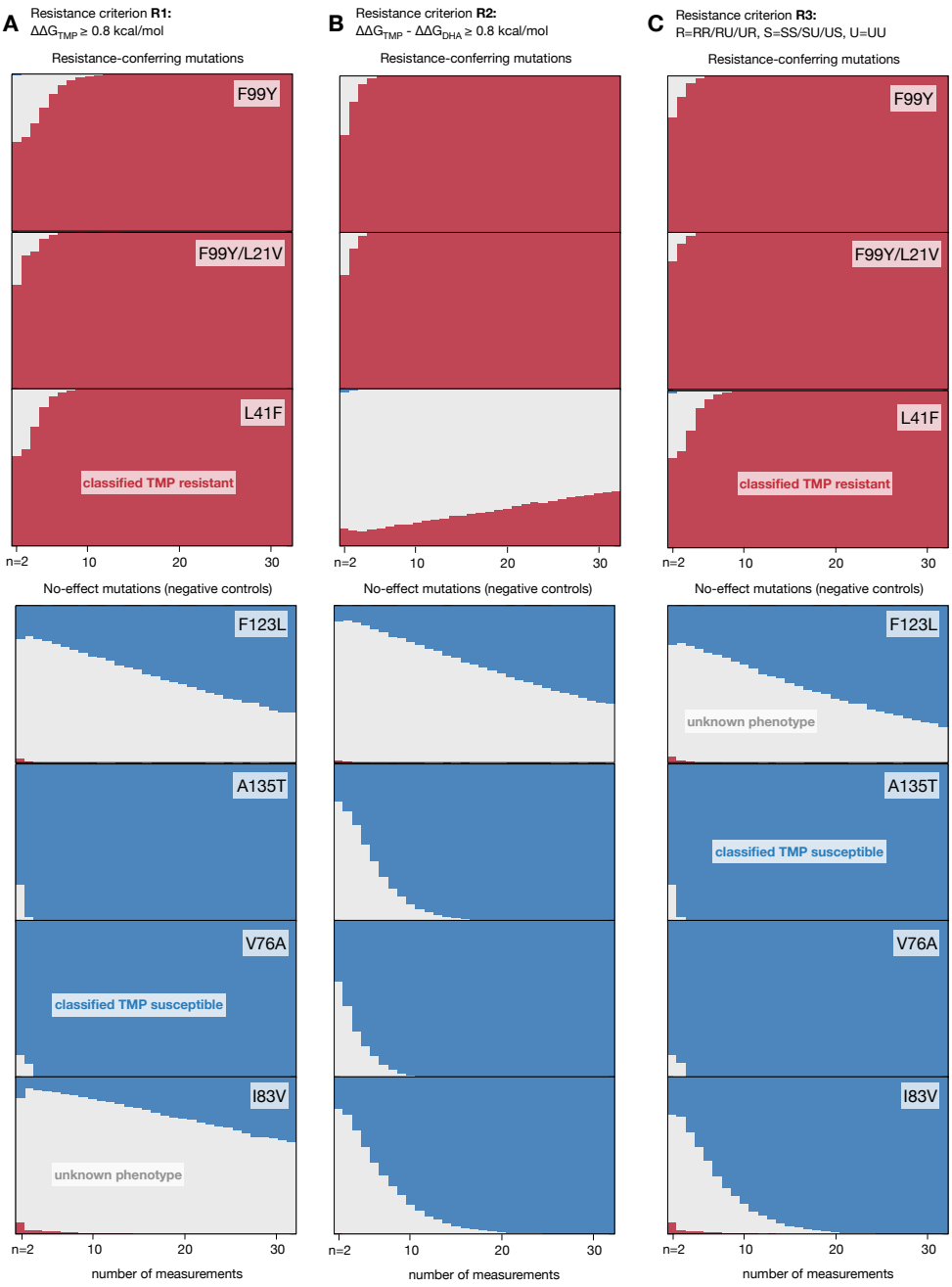

Figure S4: Related to Figure 6. Increasing the number of independent calculations improves the classification by either resistance criteria. (*A*) Using the first resistance criterion, if a small number of calculations (< 5) are run there is a small chance of susceptible mutations being classified as resistant and a moderate chance of any mutation being classified as having an unknown phenotype. As the number of calculations is increased past 10, these errors disappear and all mutations are either correctly classified (as resistant or susceptible), or an unknown result is returned. The chance of an unknown phenotype being returned falls steadily as the number of calculations increases, until at $n = 32$, we predict that five of the seven mutations, would always be correctly classified and an 'unknown' result would be returned for F123L and I83L around half the time. (*B*) The picture is similar if we apply the second resistance criterion, except that now it struggles to correctly classify the F123L and L41F mutations. Again there is a small chance of a classification error when $n < 5$, which disappears as n increases. The differences that arise from applying these two resistance criteria can be explained by considering where the mutations are found on the $\Delta\Delta G_{DHA}$ v. $\Delta\Delta G_{TMP}$ plot (Fig. 4) in relation to the lines that define both resistance criteria. (*C*) We can improve the performance slightly if we apply both resistance criteria, examine both results and allow any definitive classification ('resistant' or 'susceptible') to overrule any 'unknown' classification.

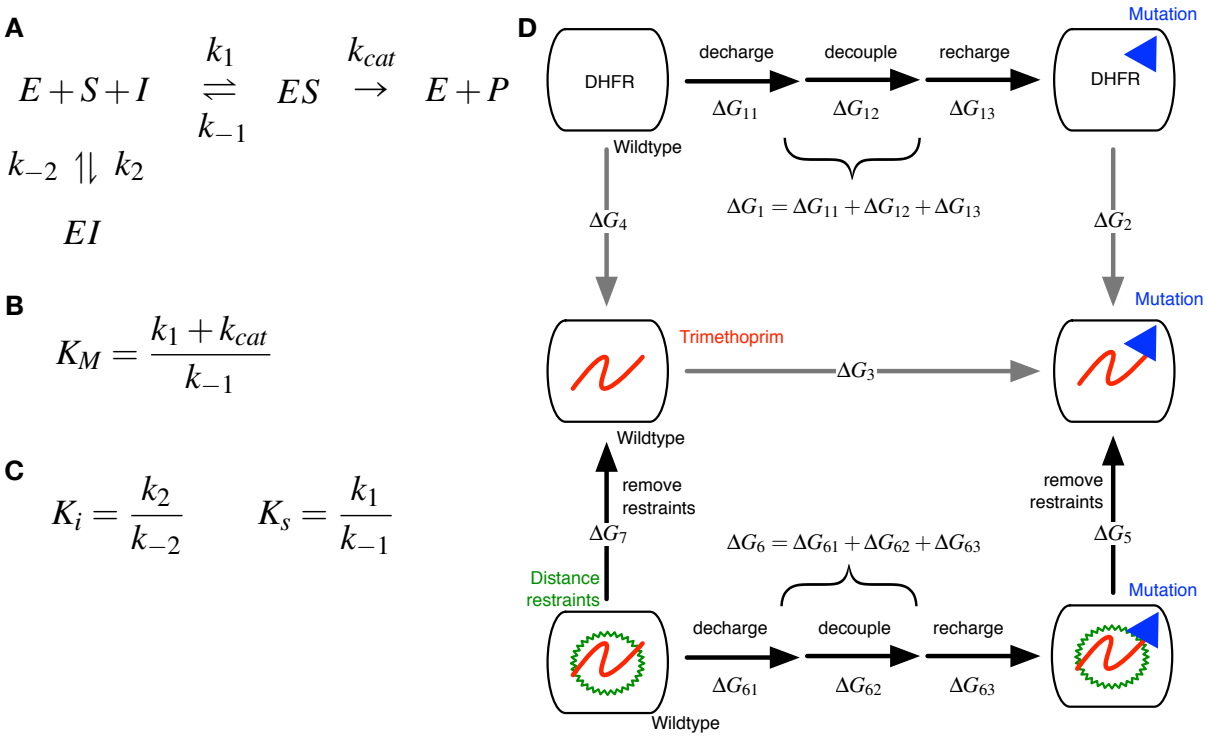

Figure S5: Related to the STAR Methods. (A) A simple kinetic scheme for the competitive inhibition of an enzyme, E, by an inhibitor, I. The enzyme binds with a substrate, S, to produce an intermediate, ES, which then reacts yielding the product, P, and the enzyme. Each step is labelled with forward and, where appropriate, reverse rate constants. (B) The Michaelis-Menten constant is defined in terms of three rate constants. (C) The dissociation constants of the inhibitor, K_i, and substrate, K_s. (D) The thermodynamic cycle used to calculate how the binding free energy of either trimethoprim or dihydrofolic acid changes ($\Delta\Delta G$) when a mutation is introduced into *S. aureus* DHFR. In the alchemical transitions (i.e. when one amino acid is transformed into another) we remove all the electrical charges on the atoms that are being perturbed, before vanishing and appearing the atoms necessary to make the mutation before finally recharging the resulting atoms. A free energy is therefore calculated separately for each step (e.g. ΔG_{11}). A soft-core van der Waals potential is used throughout. To prevent the ligand unbinding from the protein during the simulations, a restraining potential is applied. The free energy of removing this potential is calculated for both the wild-type and mutant proteins. Hence, a total of eight alchemical free energy calculations are needed for each value.